

84 872

STIC-Biot ch/ChemLib

From: Schnizer, Holly
Sent: Tuesday, January 21, 2003 4:44 PM
To: STIC-Biotech/ChemLib
Subject: seq. search for appl. no. 09/435,403

RECEIVED

JAN 22 2003

STIC-BIOTECH. (STIC)

Please search the commercial and interference databases for the following

- 1) Arg Pro Asn Tyr Ser Arg Arg Leu Pro Lys Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile
(this sequence is SEQ ID NO: 2 wherein Leu at position 3 is mutated to Asn)

Thank you.

Holly Schnizer
AU 1653
CM1-9E09
305-3722
mailbox: CM1-9B01

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 1/23
Date Completed: 1/23
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 23, 2003, 08:55:22 ; Search time 35 Seconds
(without alignments)
95.179 Million cell updates/sec

Title: SEQ2-ASNAT3
Perfect score: 136
Sequence: 1 RPhYSRRLPKGVKHLKDFPILPGEI 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	93.4	368	16 AAR73020	Human Factor-VIII
2	127	93.4	720	16 AAR74088	Factor-VIII heavy
3	127	93.4	729	16 AAR74089	Factor-VIII heavy
4	127	93.4	740	16 AAR76961	Human Factor-VIII
5	127	93.4	740	16 AAR76962	Human Factor-VIII
6	127	93.4	740	16 AAR73021	Human Factor-VIII
7	127	93.4	740	16 AAR74090	Factor-VIII heavy
8	127	93.4	740	16 AAR76982	Human factor VIII
9	127	93.4	1383	18 AAW33227	Procoagulant-activ
10	127	93.4	1383	18 AAW33228	Procoagulant-activ

11	127	93.4	1383	18	AAW33229	Procoagulant-activ
12	127	93.4	1424	9	AAP80268	Modified factor VI
13	127	93.4	1424	10	AAP91169	Sequence of 740 Ar
14	127	93.4	1424	22	AAB48842	Mutant mature huma
15	127	93.4	1425	9	AAP80267	Modified factor VI
16	127	93.4	1438	21	AAB01262	B-domain deleted f
17	127	93.4	1440	12	AAR12971	Factor VIII:SQ.
18	127	93.4	1457	19	AAW46246	Human factor VIII
19	127	93.4	1457	19	AAW44372	Human Factor VIII
20	127	93.4	1457	20	AAV21675	Beta-domain delete
21	127	93.4	1459	22	AAE10827	Human factor VIII
22	127	93.4	1459	22	AAE10832	Human factor VIII
23	127	93.4	1459	22	AAE10833	Human factor VIII
24	127	93.4	1471	18	AAW23414	Human B-domain del
25	127	93.4	1471	22	AAB67959	Amino acid sequenc
26	127	93.4	1516	9	AAP80265	Modified factor VI
27	127	93.4	1661	18	AAW18670	Factor VIII-dB695-
28	127	93.4	2098	17	AAR86863	Factor-VIII. Homo
29	127	93.4	2332	8	AAP71726	Factor VIII:c varia
30	127	93.4	2332	8	AAP71727	Factor VIII:c varia
31	127	93.4	2332	8	AAP71728	Factor VIII:c varia
32	127	93.4	2332	8	AAP71729	Factor VIII:c varia
33	127	93.4	2332	14	AAR43257	Human Factor VIII.
34	127	93.4	2332	18	AAW33222	Procoagulant-activ
35	127	93.4	2332	18	AAW33223	Procoagulant-activ
36	127	93.4	2332	18	AAW33224	Procoagulant-activ
37	127	93.4	2332	18	AAW33225	Procoagulant-activ
38	127	93.4	2332	18	AAW33226	Procoagulant-activ
39	127	93.4	2332	19	AAW53483	Human factor VIII.
40	127	93.4	2332	19	AAW44132	Homo sapiens modif
41	127	93.4	2332	20	AAV31594	Human factor VIII
42	127	93.4	2332	21	AAV57847	Human Factor VIII
43	127	93.4	2332	22	AAE10826	Human mature wild-
44	127	93.4	2332	22	AAE11200	Human factor VIII
45	127	93.4	2332	22	AAW71902	N-terminal truncat

ALIGNMENTS

RESULT 1
AAR73020
ID AAR73020 standard; peptide; 368 AA.
XX AAR73020;
XX 21-NOV-1995 (first entry)
XX Human Factor-VIII fragment.
DE Factor-VIII; blood-clotting; blood; coagulant; haemophilia-A.
XX Homo sapiens.
XX WO9513301-A.
XX 18-MAY-1995.
XX 10-NOV-1994; 94WO-DK00424.
XX 12-NOV-1993; 93DK-0001281.
XX (NOVO) NOVO-NORDISK AS.
XX Persson E;
XX WPI; 1995-194038/25.
XX Crosslinked Factor VIII polypeptide which is stable - is prepd. using bis(sulphosuccinimidy1) suberate or disuccinimidy1 suberate in the presence of polysorbate 80 to produce a coagulant with long lasting activity

PS Disclosure; Page 19; 36pp; English.

XX This is a fragment corresponding to internal AAs 373-740 of human

CC Factor-VIII which may be crosslinked resulting in

CC increased stability and retention of high activity over extended

CC periods of time after activation by thrombin. The polypeptide is

CC used to prevent or treat diseases caused by the absence or deficiency

CC of Factor-VIII in a subject such as haemophilia.

XX

SQ Sequence 368 AA;

Query Match 93.4%; Score 127; DB 16; Length 368;

Best Local Similarity 96.0%; Pred. No. 2.1e-10;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRRLPKGVKHLKDFPILPGEI 25

Db 112 RPLYSRRRLPKGVKHLKDFPILPGEI 136

RESULT 2

AAR74088

ID AAR74088 standard; protein; 720 AA.

XX

AC AAR74088;

XX

DT 04-NOV-1995 (first entry)

XX

DE Factor-VIII heavy chain N-terminal fragment.

XX

XX human; Factor VIII; heavy chain; N-terminal fragment;

KW thrombin cleavage; blood-clotting.

XX

OS Homo sapiens.

XX

PN WO9513300-A.

XX

PD 18-MAY-1995.

XX

PF 10-NOV-1994; 94WO-DK00423.

XX

PR 12-NOV-1993; 93DK-0001280.

XX

PA (NOVO) NOVO-NORDISK AS.

XX

PI Ezban Rasmussen M, Kjalke M;

XX

DR WPI; 1995-194037/25.

XX

PT Factor VIII polypeptide(s) comprising a heavy chain shorter than native

PT A1-A2 domain - are easier to produce recombinantly and retain coagulant

PT activity, may be used to treat patients who have developed antibodies to

PT C-terminal epitope(s) of Factor VIII

XX

PS Claim 3; Page 24-26; 51pp; English.

XX

CC The sequence represents N-terminal residues 1-720 of a human Factor-

CC VIII heavy chain. The sequence is shorter than the A1-A2 domain,

CC and is produced by treating a polypeptide containing the full A1-A2

CC domain of full-length Factor-VIII with a protease, e.g. thrombin.

CC The fragment has the same specific activity as full-length

CC Factor-VIII in a chromogenic assay, but activity is a factor of two

CC lower in a clotting assay, and the fragment is activated by thrombin

CC at a slower rate and to a lower level than fragments 1-740 (AAR74090),

CC 1-729 (AAR74089) and plasma Factor-VIII. The fragment may be produced

CC recombinantly to reduce production costs and improve safety, and

CC production levels and stability are higher than for the full-length

CC form. The fragment may be used to treat patients who have developed

CC antibodies against epitopes in the C-terminal part of the heavy chain.

XX

SQ Sequence 720 AA;

Query Match 93.4%; Score 127; DB 16; Length 720;

Best Local Similarity 96.0%; Pred. No. 4.3e-10;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRRLPKGVKHLKDFPILPGEI 25

Db 484 RPLYSRRRLPKGVKHLKDFPILPGEI 508

RESULT 3

AAR74089

ID AAR74089 standard; protein; 729 AA.

XX

AC AAR74089;

XX

DT 04-NOV-1995 (first entry)

XX

DE Factor-VIII heavy chain N-terminal fragment.

XX

KW human; Factor VIII; heavy chain; N-terminal fragment;

KW thrombin cleavage; blood-clotting.

XX

OS Homo sapiens.

XX

PN WO9513300-A.

XX

PD 18-MAY-1995.

XX

PF 10-NOV-1994; 94WO-DK00423.

XX

PR 12-NOV-1993; 93DK-0001280.

XX

PA (NOVO) NOVO-NORDISK AS.

XX

PI Ezban Rasmussen M, Kjalke M;

XX

DR WPI; 1995-194037/25.

XX

PT Factor VIII polypeptide(s) comprising a heavy chain shorter than native

PT A1-A2 domain - are easier to produce recombinantly and retain coagulant

PT activity, may be used to treat patients who have developed antibodies to

PT C-terminal epitope(s) of Factor VIII

XX

PS Claim 2; Page 27-29; 51pp; English.

XX

CC The sequence represents N-terminal residues 1-729 of a human Factor-

CC VIII heavy chain. The sequence is shorter than the A1-A2 domain,

CC and is produced by treating a polypeptide containing the full A1-A2

CC domain of full-length Factor-VIII with a protease, e.g. thrombin.

CC The fragment has the same coagulant specific activity as full-length

CC Factor-VIII in a chromogenic assay, and is activated by thrombin at

CC a similar rate. The fragment may be produced recombinantly to reduce

CC production costs and improve safety, and production levels and

CC stability are higher than for the full-length form. The fragment may

CC be used to treat patients who have developed antibodies against

CC epitopes in the C-terminal part of the heavy chain.

XX

SQ Sequence 729 AA;

Query Match 93.4%; Score 127; DB 16; Length 729;

Best Local Similarity 96.0%; Pred. No. 4.3e-10;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRRLPKGVKHLKDFPILPGEI 25

Db 484 RPLYSRRRLPKGVKHLKDFPILPGEI 508

RESULT 4

AAR76961

ID AAR76961 standard; protein; 740 AA.

XX

AC AAR76961;

XX

DT 09-MAR-1996 (first entry)
XX
DE Human Factor-VIII derivative.
XX
KW Factor-VIII; therapeutic; blood-clotting.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 692
FT /label= absent or Ala, Thr, Ser, Gly or Asp
FT Misc-difference 720
FT /label= absent or Gln, Ser, Thr, Val or Ala
FT Misc-difference 729
FT /label= absent or Val, Ala or Ile
XX
PN WO9518827-A1.
XX
PD 13-JUL-1995.
XX
PF 06-JAN-1995; 95WO-DK00008.
XX
PR 07-JAN-1994; 94DK-0000032.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Ezban Rasmussen M, Nicolaisen EM, Persson E;
XX WPI; 1995-255039/33.
DR
XX Novel factor VIII derivative used to treat haemophilia - and
PT comprises a functional A2 domain containing a mutation at one or
PT more Cys residues.
PT
XX Disclosure; Page 14-17; 30pp; English.
PS
XX The new Factor-VIII derivative comprises a functional A2 domain in
CC which Cys-692 is deleted or replaced with another amino acid
CC residue, preferably Ser (see AAR76962). Alternatively, Glu-720 and/or
CC Tyr-729 are deleted or substituted with various amino acids (as in
CC the Features). The new derivative has the same activity as the wild-
CC type Factor-VIII but with improved stability (the activity is
CC maintained for a longer period compared to the rapid decline of the
CC activity of wt Factor-VIII). The new derivative can be used in a
CC composition for treating diseases caused by an absence or deficiency
CC of Factor-VIII, especially haemophilia.
XX
SQ Sequence 740 AA;

Query Match 93.4%; Score 127; DB 16; Length 740;
Best Local Similarity 96.0%; Pred. NO. 4.4e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|| |||||
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 5
AAR76962
ID AAR76962 standard; protein; 740 AA.
XX
AC AAR76962;
XX
DT 09-MAR-1996 (first entry)
XX
DE Human Factor-VIII derivative.
XX
KW Factor-VIII; therapeutic; blood-clotting.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers

FT Misc-difference 692
FT /label= Cys substituted by Ser
FT Misc-difference 720
FT /label= absent or Gln, Ser, Thr, Val or Ala
FT Misc-difference 729
FT /label= absent or Val, Ala or Ile
XX
PN WO9518827-A1.
XX
PD 13-JUL-1995.
XX
PF 06-JAN-1995; 95WO-DK00008.
XX
PR 07-JAN-1994; 94DK-0000032.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Ezban Rasmussen M, Nicolaisen EM, Persson E;
XX WPI; 1995-255039/33.
DR
XX Novel factor VIII derivative used to treat haemophilia - and
PT comprises a functional A2 domain containing a mutation at one or
PT more Cys residues.
PT
XX Disclosure; Page 18-20; 30pp; English.
PS
XX The new Factor-VIII derivative comprises a functional A2 domain in
CC which Cys-692 is replaced with Ser. For other (less preferred)
CC substitutions at this site, see AAR76961. Alternatively, Glu-720 and/or
CC Tyr-729 are deleted or substituted with various amino acids (as in
CC the Features). The new derivative has the same activity as the wild-
CC type Factor-VIII but with improved stability (the activity is
CC maintained for a longer period compared to the rapid decline of the
CC activity of wt Factor-VIII). The new derivative can be used in a
CC composition for treating diseases caused by an absence or deficiency
CC of Factor-VIII, especially haemophilia.
XX
SQ Sequence 740 AA;

Query Match 93.4%; Score 127; DB 16; Length 740;
Best Local Similarity 96.0%; Pred. NO. 4.4e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|| |||||
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 6
AAR73021
ID AAR73021 standard; peptide; 740 AA.
XX
AC AAR73021;
XX
DT 21-NOV-1995 (first entry)
XX
DE Human Factor-VIII N-terminal fragment.
XX
KW Factor-VIII; blood-clotting; blood; coagulant; haemophilia-A.
XX
OS Homo sapiens.
XX
PN WO9513301-A.
XX
PD 18-MAY-1995.
XX
PF 10-NOV-1994; 94WO-DK00424.
XX
PR 12-NOV-1993; 93DK-0001281.
XX
PA (NOVO) NOVO-NORDISK AS.
XX

PI Persson E;
XX WPI; 1995-194038/25.
DR
XX
PT Crosslinked Factor VIII polypeptide which is stable - is prepd. using
PT bis(sulphosuccinimydyl) suberate or disuccinimydyl suberate in the
PT presence of polysorbate 80 to produce a coagulant with long lasting
PT activity
XX
PS Disclosure; Page 21; 36pp; English.
XX
CC This is the N-terminal fragment of human Factor-VIII which may be
CC crosslinked resulting in increased stability and retention of high
CC activity over extended periods of time after activation by thrombin.
CC The polypeptide is used to prevent or treat diseases caused by the
CC absence or deficiency of Factor-VIII in a subject such as
CC haemophilia.
XX
SQ Sequence 740 AA;

Query Match 93.4%; Score 127; DB 16; Length 740;
Best Local Similarity 96.0%; Pred. No. 4.4e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 7
AAR74090
ID AAR74090 standard; protein; 740 AA.
XX
AC AAR74090;
XX
DT 04-NOV-1995 (first entry)
XX
DE Factor-VIII heavy chain N-terminal fragment.
XX
KW human; Factor VIII; heavy chain; N-terminal fragment;
KW thrombin cleavage; blood-clotting.
XX
OS Homo sapiens.
XX
PN WO9513300-A.
XX
PD 18-MAY-1995.
XX
PF 10-NOV-1994; 94WO-DK00423.
XX
PR 12-NOV-1993; 93DK-0001280.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Ezban Rasmussen M, Kjalke M;
XX
DR WPI; 1995-194037/25.
XX
PT Factor VIII polypeptide(s) comprising a heavy chain shorter than native
PT A1-A2 domain - are easier to produce recombinantly and retain coagulant
PT activity, may be used to treat patients who have developed antibodies to
PT C-terminal epitope(s) of Factor VIII
XX
PS Disclosure; Page 30-32; 51pp; English.
XX
CC The sequence represents N-terminal residues 1-740 of a human Factor-
CC VIII heavy chain. The sequence contains entire A1 and A2 domains,
CC and truncated forms (1-720 in (AAR74088) and 1-729 in (AAR74090))
CC may be produced by treatment with a protease, e.g. thrombin. The
CC C-terminally truncated fragments have the same coagulant specific
CC activity as full-length Factor-VIII, and may be produced
CC recombinantly to reduce production costs and improve safety, giving
CC higher production levels and stability than for the full-length form.

CC The fragments may be used to treat patients who have developed
CC antibodies against epitopes in the C-terminal part of the heavy chain.
XX
SQ Sequence 740 AA;

Query Match 93.4%; Score 127; DB 16; Length 740;
Best Local Similarity 96.0%; Pred. No. 4.4e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 8
AAR76982
ID AAR76982 standard; protein; 740 AA.
XX
AC AAR76982;
XX
DT 21-FEB-1996 (first entry)
XX
DE Human factor VIII A2-domain derivative.
XX
KW Factor VIII; human; haemophilia; thrombin; protein C; plasmin;
KW serine protease; recombination; therapy; deficiency.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 720
FT /label= Gln, Ser, Thr, Val, Ala
FT Misc-difference 729
FT /label= Val, Ala, Ile
XX
PN WO9518828-A1.
XX
PD 13-JUL-1995.
XX
PF 06-JAN-1995; 95WO-DK00009.
XX
PR 07-JAN-1994; 94DK-0000031.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Ezban Rasmussen M, Nicolaisen EM, Persson E;
XX
DR WPI; 1995-255040/33.
XX
PT Novel factor VIII derivative with resistance to enzymatic cleavage -
PT and comprises a functional A2 domain where Glu720 and/or Tyr729 is
PT deleted/substituted.
XX
PS Claim 1; Page 11-14; 26pp; English.
XX
CC This sequence represents the A2 domain of a human factor VIII derivative.
CC Factor VIII is a large glycoprotein which is present in plasma at low
CC concentrations. Factor VIII is an essential part of the clotting
CC reaction in response to a wound. Factor VIII is susceptible to cleavage
CC by thrombin, activated protein C, plasmin, and other serine proteases.
CC Full length factor VIII consists of three repeats of the A-domain, a
CC B-domain and 2 repeats of the C-domain. Active factor VIII has the A1
CC domain cleaved off. Factor VIII is too unstable for use in recombinant
CC techniques. Factor VIII containing this sequence has improved stability
CC and shows resistance against enzymatic activity present in mammalian
CC cells. This means that factor VIII containing this sequence can be used
CC in recombinant techniques. A factor VIII derivative can be used for
CC treating diseases caused by an absence or deficiency of factor VIII (in
CC the same way as normal factor VIII) e.g. haemophilia. The advantage with
CC using a recombinant factor VIII also includes no need for lots of donors
CC in order to get a sufficient amount. Also, there is no long purification
CC process, and there is no risk of transmission of blood-bourne diseases
CC such as HIV.

XX SQ Sequence 740 AA;

Query Match 93.4%; Score 127; DB 16; Length 740;
Best Local Similarity 96.0%; Pred. No. 4.4e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|| |||||

Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 9
AAW33227
ID AAW33227 standard; protein; 1383 AA.

XX AC AAW33227;
XX DT 30-APR-1998 (first entry)
XX DE Procoagulant-active human factor VIII:C (FVIII) mutant protein.
XX KW pro-coagulant active factor VIII; FVIII; haemophilia A;
KW recombinant secretion; pro-coagulant activity; resistance;
KW activated protein C cleavage; APC; B domain; A2 domain; A3 domain;
KW von Willebrand factor binding site; binding affinity;
KW FVIII replacement therapy.

XX OS Synthetic.
OS Homo sapiens.

XX FH Key Location/Qualifiers
FT Region 1..346 /note= "factor VIIIA heavy chain"
FT Region 741..1383 /note= "factor VIIIA light chain"
FT Domain 1..329 /note= "A1 domain"
FT Domain 1..179 /note= "plastocyanin-like domain 1"
FT Domain 187..329 /note= "plastocyanin-like domain 2"
FT Domain 380..711 /note= "A2 domain"
FT Misc_feature 711..746 /note= "a spacer of the sequence
SFSONSRHPSTROKQFNATTIPENDIEKTDPPWF
AHRTPMPKIQNVSSDLLMLL is inserted
between domains A2 and A3"

FT Domain 380..554
FT /note= "plastocyanin-like domain 3"
FT Domain 564..711
FT /note= "plastocyanin-like domain 4"
FT Domain 746..1073
FT /note= "A3 domain"
FT Domain 1073..1221
FT /note= "C1 domain"
FT Domain 1226..1378 /note= "C2 domain"
FT Cleavage-site 372..373 /note= "by thrombin"
FT Disulfide-bond 153..179 /note= "probable"
FT Disulfide-bond 528..554 /note= "probable"
FT Misc-difference 740 /label= R740A
FT /note= "wild type Arg replaced with Ala"

XX WO9740145-A1.
XX 30-OCT-1997.

PF 24-APR-1997; 97WO-US06563.
XX
PR 15-MAY-1996; 96US-0017785.
PR 24-APR-1996; 96US-0016117.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Amano K, Kaufman RJ, Pipe SW;
XX WPI; 1997-535830/49.
DR
XX
PT Modified human pro-coagulant active factor VIII - can be
administered to haemophiliacs, i.e. factor VIII replacement therapy
XX
PS Claim 20; Page -; 57pp; English.
XX
CC The present sequence represents a novel pro-coagulant active factor
VIII (FVIII) mutant protein, comprising a deletion of the B domain and
von Willebrand factor binding site, a mutation at Arg740 and an addition
of an amino acid sequence spacer between the A2 and A3 domains. Factor
VIII, along with calcium and phospholipid, acts as a cofactor for factor
IXA, when it converts factor X to the activated form (factor XA). FVIII
is the coagulation factor deficient in the x-chromosome-linked bleeding
disorder haemophilia A. Several other mutant FVIII proteins have also
been created (see AAW33222-29). The FVIII mutant F309S (AAW33225) is
capable of recombinant secretion at higher levels than typically obtained
with wild type FVIII and retains pro-coagulant activity. The FVIII
mutant R336I (AAW33222) and R562K (AAW33223) are resistant to activated
protein C (APC) cleavage. The present FVIII mutant can form a more
stable configuration, and have an approximate 5-fold increase in
specific activity compared to purified wild type FVIII, while increasing
their binding affinity to von Willebrand factor improves their
stability. The FVIII proteins can be administered to haemophiliacs, i.e.
FVIII replacement therapy, while the nucleic acid molecule can be used
for gene therapy.
CC note: this sequence does not appear in the specification; it was created
using sequences from the given references.

XX SQ Sequence 1383 AA;

Query Match 93.4%; Score 127; DB 18; Length 1383;
Best Local Similarity 96.0%; Pred. No. 8.6e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|| |||||

Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 10
AAW33228
ID AAW33228 standard; protein; 1383 AA.
XX
AC AAW33228;
XX
DT 30-APR-1998 (first entry)
XX
DE Procoagulant-active human factor VIII:C (FVIII) mutant protein.
XX
KW Pro-coagulant active factor VIII; FVIII; haemophilia A;
KW recombinant secretion; pro-coagulant activity; resistance;
KW activated protein C cleavage; APC; B domain; A2 domain; A3 domain;
KW von Willebrand factor binding site; binding affinity;
KW FVIII replacement therapy.

XX OS Synthetic.
OS Homo sapiens.

XX FH Key Location/Qualifiers
FT Region 1..346 /note= "factor VIIIA heavy chain"
FT Region 741..1383 /note= "factor VIIIA light chain"
FT Region 741..1383 /note= "factor VIIIA light chain"

FT Domain 1..329 /note= "A1 domain"
FT Domain 1..179
FT /note= "plastocyanin-like domain 1"
FT 187..329
FT /note= "plastocyanin-like domain 2"
FT 380..711
FT /note= "A2 domain"
FT 711..746
FT /note= "a spacer of the sequence
FT SFQNSRHPSTROKQFNATTPENDIEKTDPPWF
FT AHRTPMPKIQNVSSDLLMLL is inserted
FT between domains A2 and A3"
FT
FT Domain 380..554
FT /note= "plastocyanin-like domain 3"
FT 564..711
FT /note= "plastocyanin-like domain 4"
FT 746..1073
FT /note= "A3 domain"
FT 1073..1221
FT /note= "C1 domain"
FT 1226..1378
FT /note= "C2 domain"
FT 372..373
FT /note= "by thrombin"
FT Disulfide-bond 153..179
FT /note= "probable"
FT Disulfide-bond 528..554
FT /note= "probable"
FT Misc-difference 336
FT /label= R336I
FT /note= "wild type Arg replaced with Ile"
FT Misc-difference 562
FT /label= R562K
FT /note= "wild type Arg replaced with Lys"
FT Misc-difference 740
FT /label= R740A
FT /note= "wild type Arg replaced with Ala"
XX
PN WO9740145-A1.
XX
PD 30-OCT-1997.
PF 24-APR-1997; 97WO-US06563.
XX
PR 15-MAY-1996; 96US-0017785.
PR 24-APR-1996; 96US-0016117.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Amano K, Kaufman RJ, Pipe SW;
DR WPI; 1997-535830/49.
XX
PT Modified human pro-coagulant active factor VIII - can be
PT administered to haemophiliacs, i.e. factor VIII replacement therapy
XX
PS Claim 18; Page -: 57pp; English.
XX
CC The present sequence represents a novel pro-coagulant active factor
CC VIII (FVIII) mutant protein, comprising a deletion of the B domain and
CC von Willebrand factor binding site, mutations R336I, R562K and R740A and
CC an addition of an amino acid sequence spacer between the A2 and A3
CC domains. Factor VIII, along with calcium and phospholipid, acts as a
CC cofactor for factor IXA, when it converts factor X to the activated form
CC (factor XA). FVIII is the coagulation factor deficient in the
CC X-chromosome-linked bleeding disorder haemophilia A. Several other
CC mutant FVIII proteins have also been created (see AAW33222-29). The FVIII
CC mutant F309S (AAW33225) is capable of recombinant secretion at higher
CC levels than typically obtained with wild type FVIII and retains
CC pro-coagulant activity. The FVIII mutant R336I (AAW33222) and R562K
CC (AAW33223) are resistant to activated protein C (APC) cleavage. The FVIII
CC mutant comprising a deletion of the B domain and von Willebrand factor

CC binding site, a mutation at Arg740 and an addition of an amino acid
CC sequence spacer between the A2 and A3 domains can form a more stable
CC configuration, and have an approximate 5-fold increase in specific
CC activity compared to purified wild type FVIII, while increasing their
CC binding affinity to von Willebrand factor improves their stability.
CC The FVIII proteins can be administered to haemophiliacs, i.e. FVIII
CC replacement therapy, while the nucleic acid molecule can be used for
CC gene therapy.
CC note: this sequence does not appear in the specification; it was created
CC using sequences from the given references.
XX
SQ Sequence 1383 AA;
Query Match 93.4%; Score 127; DB 18; Length 1383;
Best Local Similarity 96.0%; Pred. No. 8.6e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508
RESULT 11
AAW33229
ID AAW33229 standard; protein; 1383 AA.
XX
AC AAW33229;
XX 30-APR-1998 (first entry)
DT XX
DE Procoagulant-active human factor VIII:C (FVIII) mutant protein.
XX
KW Pro-coagulant active factor VIII; FVIII; haemophilia A;
KW recombinant secretion; pro-coagulant activity; resistance;
KW activated protein C cleavage; APC; B domain; A2 domain; A3 domain;
KW von Willebrand factor binding site; binding affinity;
KW FVIII replacement therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1..346
FT /note= "factor VIIIA heavy chain"
FT 741..1383
FT /note= "factor VIIIA light chain"
FT Domain 1..329
FT /note= "A1 domain"
FT Domain 1..179
FT /note= "plastocyanin-like domain 1"
FT Domain 187..329
FT /note= "plastocyanin-like domain 2"
FT Domain 380..711
FT /note= "A2 domain"
FT Misc_feature 711..746
FT /note= "a spacer of the sequence
FT SFQNSRHPSTROKQFNATTPENDIEKTDPPWF
FT AHRTPMPKIQNVSSDLLMLL is inserted
FT between domains A2 and A3"
FT
FT Domain 380..554
FT /note= "plastocyanin-like domain 3"
FT Domain 564..711
FT /note= "plastocyanin-like domain 4"
FT Domain 746..1073
FT /note= "A3 domain"
FT Domain 1073..1221
FT /note= "C1 domain"
FT Domain 1226..1378
FT /note= "C2 domain"
FT Cleavage-site 372..373
FT /note= "by thrombin"
FT Disulfide-bond 153..179
FT /note= "probable"

XX WPI; 1989-078467/11.
 DR N-PSDB; AAN90654.
 XX
 PT Prodn. of recombinant human Factor-VIII-C -
 PT using animal cells transformed with a vector contg. the gene for
 PT Factor VIII:C and a promoter
 XX
 PS Fig 1(1) - 1(13); ; 32pp; English.
 XX
 CC Arg-740 of the carboxyl terminus of the H chain is
 CC directly bonded by a peptide bond to Glu-1649 of the amino terminus of
 CC L chain. A pref. expression vector used to transform animal cell so
 CC that they produce human Factor VIII:Cis plasmid Ad.RE.neo.
 CC The expression vector has at least one promoter upstream of AAN90654.
 CC The transformants can constantly and continuously produce human Factor
 CC VIII:C in high yield on a commercial scale. The human Factor VIII:C so
 CC produced is considered to corresp. to the smallest species of active and
 CC intact Factor VIII:C molecules in the human blood plasma. It is useful
 CC for treating haemophilia A patients.
 XX
 SQ Sequence 1424 AA;
 Query Match 93.4%; Score 127; DB 10; Length 1424;
 Best Local Similarity 96.0%; Pred. No. 8.8e-10;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
 DB 484 RPLYSRRLPKGVKHLKDFPILPGEI 508
 RESULT 14
 AAB48842
 ID AAB48842 standard; protein; 1424 AA.
 XX
 AC AAB48842;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Mutant mature human factor VIII, SEQ ID NO:5.
 XX
 KW Factor VIII; human; B domain; LRP-mediated plasma clearance;
 KW receptor-dependent clearance; receptor-independent clearance;
 KW half-life; haemophilia; mutant; mutein.
 XX
 OS Homo sapiens.
 XX
 PN WO200071714-A2.
 XX
 PD 30-NOV-2000.
 XX
 PF 24-MAY-2000; 2000WO-US14111.
 XX
 PR 24-MAY-1999; 99US-0135847.
 XX
 PA (AMNA-) AMERICAN NAT RED CROSS.
 PI Saenko EL, Strickland DK;
 XX
 DR WPI; 2001-025163/03.
 XX
 PT Factor VIII mutants having increased half-life useful for treating
 PT hemophilia, comprise one or more amino acid substitutions in the A2
 PT and/or C2 domain of factor VIII -
 XX
 PS Claim 9; Fig 2A-B; 121pp; English.
 XX
 CC The invention relates to human factor VIII mutants comprising an amino
 CC acid substitution at one or more positions in the A2 domain and/or an
 CC amino acid substitution at one or more positions in the C2 domain.
 CC The invention also encompasses a factor VIII mutant which lacks a B
 CC domain (AAB48842). The factor VIII mutants have an increased half-life

CC in the bloodstream. The A2 domain mutants exhibit reduced LRP-dependent
 CC (receptor-dependent) clearance of factor VIII, while C2 domain mutants
 CC have reduced receptor-independent clearance. The invention also relates
 CC to a method of using RAP (receptor associated protein), a protein which
 CC inhibits LRP (low density lipoprotein related protein)-mediated ligand
 CC internalisation, to increase the half-life of factor VIII. The mutant
 CC factor VIII proteins, and nucleotides encoding them, are useful
 CC for treating haemophilia. RAP, LRP-binding RAP mutants or fragments, and
 CC nucleic acids encoding them may also be used in the treatment of
 CC haemophilia, in combination with a mutant factor VIII protein or DNA of
 CC the invention. The invention provides means of increasing the half-life
 CC of factor VIII by reducing its clearance from plasma. The present
 CC sequence represents a mutant mature human factor VIII which lacks a B
 CC domain.
 XX
 SQ Sequence 1424 AA;
 Query Match 93.4%; Score 127; DB 22; Length 1424;
 Best Local Similarity 96.0%; Pred. No. 8.8e-10;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
 DB 484 RPLYSRRLPKGVKHLKDFPILPGEI 508
 RESULT 15
 AAP80267
 ID AAP80267 standard; protein; 1425 AA.
 XX
 AC AAP80267;
 XX
 DT 10-OCT-1990 (first entry)
 XX
 DE Modified factor VIII:C sequence with the R740-E1649 deletion.
 XX
 KW Modified factor VIII:C; maturation polypeptide; haemophilia;
 KW blood coagulation; RE deletion.
 XX
 OS Homo sapiens.
 XX
 PN WO8800831-A.
 XX
 PD 11-FEB-1988.
 XX
 PF 31-JUL-1987; 87WO-US01814.
 XX
 PR 01-AUG-1986; 86US-0893375.
 XX
 PA (BIOJ) BIOGEN NV (PASE/).
 XX
 PI Pasek MP;
 XX
 DR WPI; 1988-049866/07.
 DR N-PSDB; AAN80446.
 XX
 PT New DNA sequences encoding modified factor VIII:C - with deletion of DNA
 PT encoding maturation polypeptide, useful for high yield transformation.
 XX
 PS Claim 3; Page 57-58-59-60; 97pp; English.
 XX
 CC The entire sequence encoding the maturation polypeptide of
 CC factor VIII:C is deleted, i.e. Arg 740-Glu 1649.
 CC The full length Factor VIII:C cDNA has two changes with respect to the
 CC published sequence (EPO application 160457):
 CC CTG to CTA at Leu 242 and TTC to CTC change at amino acid residue 1880
 CC (Phe to Leu). The product is produced in approx. 20 times higher
 CC yields than previous recombinant produced factor VIII:C and are more
 CC easily purified. The peptide is used for treating haemophilia A, both
 CC acute and prolonged bleeding.
 CC See also AAN80444 and AAN80447.
 XX
 SQ Sequence 1425 AA;

Query Match 93.4%; Score 127; DB 9; Length 1425;
Best Local Similarity 96.0%; Pred. No. 8.8e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|||
Db 485 RPLYSRRLPKGVKHLKDFPILPGEI 509

Search completed: January 23, 2003, 09:00:54
Job time : 38 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - protein search, using sw model
Run on: January 23, 2003, 09:01:18 ; Search time 55 Seconds
(without alignments)
43.697 Million cell updates/sec

Title: SEQ2-ASNAT3
Perfect score: 136
Sequence: 1 RPNYSRRLPKGVKHLKDFPILPGEI 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	127	93.4	2351 1	EZHU coagulation factor
2	94	69.1	2319 2	A47004 coagulation factor
3	75	55.1	2133 2	T42763 coagulation factor
4	51	37.5	1607 2	T04583 TMV resistance pro
5	50	36.8	958 2	A82583 conserved hypothet
6	49	36.0	412 2	G71070 hypothetical prote
7	48.5	35.7	400 2	JC4591 alpha-1,3 fucosylt
8	48.5	35.7	405 2	B36340 alpha(1,3)-fucosyl
9	48.5	35.7	433 2	A57596 alpha-1,3-fucosylt
10	48.5	35.7	1253 2	T45787 disease resistance
11	48	35.3	156 2	B83164 conserved hypothet
12	48	35.3	264 2	AC2515 hypothetical prote
13	48	35.3	600 2	B46642 DNA-directed DNA p
14	48	35.3	1690 2	T40847 probable rRNA biog
15	47	34.6	444 2	S54011 l-aminocyclopropan
16	46	33.8	308 2	S67657 hypothetical prote
17	46	33.8	348 2	G71681 hypothetical prote
18	46	33.8	467 2	T38353 serine hydroxymeth
19	46	33.8	480 2	A99164 hypothetical prote
20	46	33.8	519 2	S77572 oligopeptide trans
21	46	33.8	885 2	T38387 hypothetical prote
22	46	33.8	1001 2	C88779 protein T20D3.9 [i
23	46	33.8	1023 2	A59431 KIAA0013 protein [
24	46	33.8	1038 2	T25033 hypothetical prote
25	45.5	33.5	161 2	F75633 hypothetical prote
26	45.5	33.5	246 2	H82553 3-demethylubiquino
27	45.5	33.5	554 2	B90679 3-(3-hydroxyphenyl
28	45.5	33.5	554 2	F85529 3-(3-hydroxyphenyl
29	45.5	33.5	554 2	C64762 probable monooxyge

ALIGNMENTS

RESULT 1

EZHU
coagulation factor VIII precursor [validated] - human
N;Alternate names: antihemophilic factor A; coagulation factor VIIIC; procoagulant co
C;Species: Homo sapiens (man)
C;Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 08-Dec-2000
C;Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S66445;
R;Gitschier, J.; Wood, W.I.
Hum. Mol. Genet. 1, 199-200, 1992
A;Title: Sequence of the exon-containing regions of the human factor VIII gene.
A;Reference number: I54318; MUID:93265012; PMID:1303178
A;Accession: I54318
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1921,'S',1923-2351 <RES>
A;Cross-references: GB:M88648; NID:g182381; PIDN:AAA52420.1; PID:g182383
R;Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschier, J.; Keyt, B.; Seeb
Nature 312, 330-337, 1984
A;Title: Expression of active human factor VIII from recombinant DNA clones.
A;Reference number: A00525; MUID:85061548; PMID:6438526
A;Accession: A00525
A;Molecule type: mRNA
A;Residues: 1-2351 <WOO>
A;Cross-references: EMBL:X01165; EMBL:X01166; EMBL:X01179
R;Toole, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.L.; Pittman, D.D
s, D.N.; Hewick, R.M.
Nature 312, 342-347, 1984
A;Title: Molecular cloning of a cDNA encoding human antihemophilic factor.
A;Reference number: I58059; MUID:85061550; PMID:6438528
A;Accession: I58059
A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-74,'V',76-1259,'E',1261-2351 <RE2>
A;Cross-references: GB:K01740; NID:g182802; PIDN:AAA52484.1; PID:g182803
R;Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.; K
.B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen, J.;
DNA 4, 333-349, 1985
A;Title: Characterization of the polypeptide composition of human factor VIII:C and t
A;Reference number: A23584; MUID:86081164; PMID:3935400
A;Accession: A23584
A;Molecule type: mRNA
A;Residues: 1-2351 <TRU>
A;Cross-references: GB:M14113; NID:g182817; PIDN:AAA52485.1; PID:g182818
R;Eaton, D.; Rodriguez, H.; Vehar, G.A.
Biochemistry 25, 505-512, 1986
A;Title: Proteolytic processing of human factor VIII. Correlation of specific cleavag
ity.
A;Reference number: A26174; MUID:86159740; PMID:3082357
A;Accession: A26174
A;Molecule type: protein
A;Residues: 20-36;392-399,'X',401-402;1668-1678;1709-1722,'D',1723-1725;1741-1755 <EA
R;Pittman, D.D.; Wang, J.H.; Kaufman, R.J.

hypothetical prote
hypothetical prote
UDP-N-acetylglucos
hypothetical prote
collagen alpha 4(I
collagen alpha 6(I
gag-pol polyprotei
hypothetical prote
lipoprotein (impor
IMP dehydrogenase
phenylalanine-tRNA
stable tubule only
hypothetical prote
hypothetical prote
ribosomal protein
hypothetical prote

Query Match 55.1%; Score 75; DB 2; Length 2133;
Best Local Similarity 83.3%; Pred. No. 0.014;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 RLPKGVKHLKDFPILPGE 24
||| ||||| |||||
Db 509 RLLKGWKHLKDMPIIPGE 526

RESULT 4
T04583
TMV resistance protein N homolog F23E13.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 17-Nov-2000
C:Accession: T04583; T05507
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse, T.
submitted to the Protein Sequence Database, March 1998
A:Reference number: Z15378
A:Accession: T04583
A:Molecule type: DNA
A:Residues: 1-1607 <BEV>
A:Cross-references: EMBL:AL022141
A:Experimental source: cultivar Columbia; BAC clone F23E13
R:Bevan, M.; Wedler, H.; Wambutt, R.; Hobeisel, J.; Mewes, H.W.; Mayer, K.F.X.; Schueller
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15418
A:Accession: T05507
A:Molecule type: DNA
A:Residues: 1448-1607 <BE2>
A:Cross-references: EMBL:AL022373
A:Experimental source: cultivar Columbia; BAC clone T19K4
C:Genetics:
A:Map position: 4
A:Introns: 193/2; 238/2; 556/2; 930/3; 1029/3; 1287/3
A:Note: F23E13.30; T19K4.270

Query Match 37.5%; Score 51; DB 2; Length 1607;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 NYSRRLPKGVKHLKD 17
:| | ||||| | | |
Db 973 HYSRLPKGLKFLPD 987

RESULT 5
A82583
conserved hypothetical protein XF2237 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: A82583
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: A82583
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-958 <SIM>
A:Cross-references: GB:AE004036; GB:AE003849; NID:g9107384; PIDN:AAF85036.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2237

Query Match 36.8%; Score 50; DB 2; Length 958;
Best Local Similarity 47.1%; Pred. No. 34;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 6 RRLPKGVKHLKDFPILP 22
||:| |: | |:
Db 124 RRIPKGIQETPDITIP 140

RESULT 6
G71070
hypothetical protein PH1259 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C:Accession: G71070
R:Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophili
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: G71070
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-412 <KAW>
A:Cross-references: GB:AP000005; NID:g3236132; PIDN:BAA30361.1; PID:g3257678
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBa
C:Genetics:
A:Gene: PH1259
C:Superfamily: hypothetical protein HI0333

Query Match 36.0%; Score 49; DB 2; Length 412;
Best Local Similarity 58.8%; Pred. No. 19;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 PNYSRRLPKGVKHLKDF 18
| : | | | : | | | |
Db 140 PIFSERTPKYLKALKDF 156

RESULT 7
JC4591
alpha-1,3 fucosyltransferase (EC 2.4.1.-) - mouse
C:Species: Mus musculus (house mouse)
C:Date: 16-Apr-1996 #sequence_revision 24-May-1996 #text_change 24-Nov-1999
C:Accession: JC4591
R:Ozawa, M.; Muramatsu, T.
J. Biochem. 119, 302-308, 1996
A:Title: Molecular cloning and expression of a mouse alpha-1,3 fucosyltransferase gen
A:Reference number: JC4591; MUID:97037075; PMID:8882722
A:Accession: JC4591
A:Molecule type: mRNA
A:Residues: 1-400 <OZA>
A:Cross-references: DDBJ:D63379
A:Experimental source: Embryonal carcinoma F9 cells
C:Superfamily: galactoside 3(4)-L-fucosyltransferase
C:Keywords: glycoprotein; glycosyltransferase; hexosyltransferase; transmembrane prot
F;1-23/Domain: intracellular #status predicted <INT>
F;24-49/Domain: transmembrane #status predicted <TRM>
F;84,185/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 35.7%; Score 48.5; DB 2; Length 400;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 1 RPNYSRRLPKGV-KHLKDFP 19

```
Db 314 RANVERFVPRGAFIHVDFFP 333
      | | | | | : | : | | | | | : | | | |
RESULT 8
B36340
alpha(1,3)-fucosyltransferase (EC 2.4.1.1-) 4 precursor [validated] - human
N;Alternate names: CD15; ELAM-1 ligand fucosyltransferase (ELFT); FCT3A; FUC-TIV; myeloid
C;Species: Homo sapiens (man)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: B36340; A36340; A40976; A41202
R;Goelz, S.E.; Hession, C.; Goff, D.; Griffiths, B.; Tizard, R.; Newman, B.; Chi-Rosso;
Cell 63, 1349-1356, 1990
A;Title: ELFT: a gene that directs the expression of an ELAM-1 ligand.
A;Reference number: A36340; MUID:91084863; PMID:1702034
A;Accession: B36340
A;Molecule type: mRNA
A;Residues: 1-405 <GOE1>
A;Cross-references: GB:M58596; NID:g182068; PIDN:AAA63172.1; PID:g182069
A;Accession: A36340
A;Molecule type: mRNA
A;Residues: 'MRLWGAARKPSGAGWEKEWAEAPQAPGWSGRLGPGR', 'SGRKGRAVPGWASWPAHLAALARPRLGGAGQ
A;Cross-references: GB:M58597; NID:g182070; PIDN:AAA63173.1; PID:g182071
A;Note: the codon used as an initiator for this translation is not in a good context for
R;Lowe, J.B.; Kukowska-Latallo, J.F.; Nair, R.P.; Larsen, R.D.; Marks, R.M.; Macher, B.A
J. Biol. Chem. 266, 17467-17477, 1991
A;Title: Molecular cloning of a human fucosyltransferase gene that determines expression
A;Reference number: A40976; MUID:91373370; PMID:1716630
A;Accession: A40976
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-86,'p',88-405 <LOW>
A;Cross-references: GB:M65030; NID:g182791; PIDN:AAA92977.1; PID:g1236720
R;Kumar, R.; Potvin, B.; Muller, W.A.; Stanley, P.
J. Biol. Chem. 266, 21777-21783, 1991
A;Title: Cloning of a human alpha(1,3)-fucosyltransferase gene that encodes ELFT but doe
A;Reference number: A41202; MUID:92042084; PMID:1718983
A;Accession: A41202
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-240,'d',242-400 <KUM>
A;Cross-references: GB:S65161; NID:g239005; PIDN:AAB20349.1; PID:g239006
C;Genetics:
A;Gene: GDB:FUT4; CD15; FCT3A; FUC-TIV
A;Cross-references: GDB:131563; OMIM:104230
A;Map position: 11q21-11q21
C;Superfamily: galactoside 3(4)-L-fucosyltransferase
C;Keywords: glycoprotein; glycosyltransferase; hexosyltransferase
F;1-48/Domain: signal sequence #status predicted <SIG>
F;49-405/Product: alpha(1,3)-fucosyltransferase 4 #status predicted <MAT>
F;91,190/Binding site: carbohydrate (Asn) (covalent) #status predicted

      Query Match 35.7%; Score 48.5; DB 2; Length 405;
      Best Local Similarity 50.0%; Pred. No. 23;
      Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

Qy 1 RPNYSRRLPKGV-KHLKDFP 19
      | | | | | : | : | | | | | : | | | |
Db 319 RANVERFVPRGAFIHVDFFP 338
      | | | | | : | : | | | | | : | | | |
RESULT 9
A57596
alpha-1,3-fucosyltransferase Fuc-TIV (EC 2.4.1.1-) - mouse
N;Alternate names: ELAM-1 ligand fucosyltransferase homolog
C;Species: Mus musculus (house mouse)
C;Date: 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change 11-Jan-2000
C;Accession: A57596
R;Gersten, K.M.; Natsuka, S.; Trinchera, M.; Petryniak, B.; Kelly, R.J.; Hiraiwa, N.; Je
J. Biol. Chem. 270, 25047-25056, 1995
A;Title: Molecular cloning, expression, chromosomal assignment, and tissue-specific expn
erases.
A;Reference number: A57596; MUID:96027607; PMID:7559635

A;Accession: A57596
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-433 <GER>
A;Cross-references: GB:U33457; NID:g1039426; PIDN:AAC52269.1; PID:g1039427
C;Superfamily: galactoside 3(4)-L-fucosyltransferase
C;Keywords: glycosyltransferase; hexosyltransferase

      Query Match 35.7%; Score 48.5; DB 2; Length 433;
      Best Local Similarity 50.0%; Pred. No. 24;
      Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

Qy 1 RPNYSRRLPKGV-KHLKDFP 19
      | | | | | : | : | | | | | : | | | |
Db 347 RANVERFVPRGAFIHVDFFP 366
      | | | | | : | : | | | | | : | | | |
RESULT 10
T45787
disease resistance-like protein - Arabidopsis thaliana
N;Alternate names: protein F26013.200
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T45787
R;Delseny, M.; Berger, C.; Cooke, R.; Grellet, F.; Laudie, M.; Mewes, H.W.; Lemcke, K
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z23013
A;Accession: T45787
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1253 <DEL>
A;Cross-references: EMBL:AL133452
A;Experimental source: cultivar Columbia; BAC clone F26013
C;Genetics:
A;Map position: 3
A;Introns: 138/2; 490/3; 589/3; 1047/3
A;Note: F26013.200

      Query Match 35.7%; Score 48.5; DB 2; Length 1253;
      Best Local Similarity 39.3%; Pred. No. 76;
      Matches 11; Conservative 4; Mismatches 4; Indels 9; Gaps 1;

Qy 3 NYSRRLPKGVK-----HLKDFPIL 21
      | | | | | | | | | | | | | | | | | | | |
Db 533 HYALHLPKGVKSLPEELRLHLHWEHPILL 560
      | | | | | | | | | | | | | | | | | | | |
RESULT 11
B83164
conserved hypothetical protein PA3847 [imported] - Pseudomonas aeruginosa (strain PAO
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: B83164
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: B83164
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-156 <STO>
A;Cross-references: GB:AE004802; GB:AE004091; NID:g9950021; PIDN:AAG07234.1; GSPDB:GN
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA3847

      Query Match 35.3%; Score 48; DB 2; Length 156;
      Best Local Similarity 36.4%; Pred. No. 9.7;
      Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPKGVKHLKDFPILP 22
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 23, 2003, 08:55:57 ; Search time 11 Seconds
(without alignments)
94.264 Million cell updates/sec

Title: SEQ2-ASNAT3
Perfect score: 136
Sequence: 1 RPNYSRRLPKGVKHLKDFILPGEI_25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	93.4	2351	1 FA8_HUMAN	P00451 homo sapien
2	94	69.1	2319	1 FA8_MOUSE	Q06194 mus musculu
3	75	55.1	2133	1 FA8_PIG	P12263 sus scrofa
4	48.5	35.7	405	1 FUT4_HUMAN	P22083 homo sapien
5	48.5	35.7	433	1 FUT4_MOUSE	Q11127 mus musculu
6	48	35.3	495	1 DPO2_RAT	O89043 rattus norv
7	48	35.3	600	1 DPO2_MOUSE	P33611 mus musculu
8	46	33.8	467	1 GLYD_SCHPO	O13972 schizosacch
9	45.5	33.5	246	1 UBIG_XYLFA	Q9pam5 xylella fas
10	45.5	33.5	554	1 MHPA_ECOLI	P77397 escherichia
11	45	33.1	438	1 MURA_SYNY3	Q55673 synechocyst
12	45	33.1	1032	1 ITA4_XENLA	Q91687 xenopus lae
13	45	33.1	1678	1 CA64_HUMAN	Q14031 homo sapien
14	45	33.1	1690	1 CA44_HUMAN	P53420 homo sapien
15	45	33.1	2054	1 YCF2_PINTH	P41653 pinus thunb
16	44.5	32.7	788	1 SYFB_THEMEA	Q9wzs9 thermotoga
17	44	32.4	130	1 RR11_MARPO	P06364 marchantia
18	44	32.4	300	1 ERA_MYCTU	O05834 mycobacteri
19	44	32.4	356	1 CCDA_CAEEL	P35800 caenorhabdi
20	44	32.4	385	1 VATC_METTH	O27038 methanobact
21	44	32.4	1143	1 CALI_HUMAN	Q14993 homo sapien
22	43.5	32.0	251	1 RPC8_YEAST	P17890 saccharomyc
23	43.5	32.0	352	1 AROB_SULTO	Q96y96 sulfolobus
24	43.5	32.0	649	1 VATI_CHLTR	O84307 chlamydia t
25	43	31.6	146	1 YZ37_AQUAE	O66425 aquifex aeo
26	43	31.6	156	1 RS10_LJMRU	O77302 lumbricus r
27	43	31.6	165	1 RS10_XENLA	Q07254 xenopus lae
28	43	31.6	166	1 RS10_TCTPU	Q90yr4 ictalurus p
29	43	31.6	227	1 FLPA_PYRHO	O57811 pyrococcus
30	43	31.6	441	1 CYSP_THEAN	P25781 theileria a
31	43	31.6	529	1 T1M_SALPO	P07989 salmonella
32	43	31.6	529	1 T1M_SALTY	P40813 salmonella
33	43	31.6	579	1 ILVB_SPIPL	P27868 spirulina p

RESULT 1				
FA8_HUMAN				
ID	FA8_HUMAN	STANDARD;	PRT;	2351 AA.
AC	P00451;			
DT	21-JUL-1986	(Rel. 01, Created)		
DT	21-JUL-1986	(Rel. 01, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Coagulation factor VIII precursor (Procoagulant component)			
DE	(Antihemophilic factor) (AHF).			
GN	F8 OR F8C.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86081164; PubMed=3935400;			
RA	Truett M.A., Blacher R., Burke R.L., Caput D., Chu C., Dina D.,			
RA	Hartog K., Kuo C.H., Masiaz F.R., Merryweather J.P., Najarian R.,			
RA	Pachl C., Potter S.J., Puma J., Quiroga M., Rall L.B., Randolph A.,			
RA	Urdea M.S., Valenzuela P., Dahl H.-H.M., Favalaro J., Hansen J.,			
RA	Nordfang O., Ezban M.;			
RT	*Characterization of the polypeptide composition of human factor			
RT	VIII:C and the nucleotide sequence and expression of the human kidney			
RT	cDNA.";			
RL	DNA 4:333-349(1985).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=85061548; PubMed=6438526;			
RA	Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschier J.,			
RA	Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wion K.L.,			
RA	Delwart E., Tuddenham E.G.D., Vehar G.A., Lawn R.M.;			
RT	*Expression of active human factor VIII from recombinant DNA clones.;"			
RL	Nature 312:330-337(1984).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=85061550; PubMed=6438528;			
RA	Toole J.J., Knopf J.L., Wozney J.M., Sultzman L.A., Buecker J.L.,			
RA	Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C.,			
RA	Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Fass D.N.,			
RA	Hewick R.M.;			
RT	*Molecular cloning of a cDNA encoding human antihaemophilic factor.;"			
RL	Nature 312:342-347(1984).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93265012; PubMed=1303178;			
RA	Gitschier J., Wood W.I.;			
RT	*Sequence of the exon-containing regions of the human factor VIII			
RT	gene.;"			
RL	Hum. Mol. Genet. 1:199-200(1992).			
RN	[5]			
RP	SEQUENCE OF 2064-2070 FROM N.A.			
RA	de Water N.S., Williams R., Browett P.J.;			
RL	Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.			
RN	[6]			
RP	SULFATION OF TYR-1699.			

Q9uy55 pyrococcus
P10964 saccharomyc
P55200 mus musculu
P70066 xiphophorus
P51864 homo sapien
Q9pk88 chlamydia m
P00575 escherichia
P06173 salmonella
P06365 nicotiana t
O05132 neisseria g
Q9jvw1 neisseria m
O84817 chlamydia t

ALIGNMENTS

RX MEDLINE=91093266; PubMed=1898735;
RA Leyte A., van Schijndel H.B., Niehrs C., Huttner W.B., Verbeet M.P.,
RA Mertens K., van Mourik J.A.;
RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is
RT essential for the interaction of factor VIII with von Willebrand
RT factor.";
RL J. Biol. Chem. 266:740-746(1991).
RN [7]
RP SULFATION.
RX MEDLINE=92207952; PubMed=1554716;
RA Pittman D.D., Wang J.H., Kaufman R.J.;
RA "Identification and functional importance of tyrosine sulfate
RT residues within recombinant factor VIII.";
RL Biochemistry 31:3315-3325(1992).
RN [8]
RP STRUCTURE BY NMR OF 2322-2343.
RX MEDLINE=95200924; PubMed=7893714;
RA Gilbert G.E., Baleja J.D.;
RT "Membrane-binding peptide from the C2 domain of factor VIII forms an
RT amphipathic structure as determined by NMR spectroscopy.";
RL Biochemistry 34:3022-3031(1995).
RN [9]
RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE=91221499; PubMed=1902642;
RA Gitschier J.;
RT "The molecular basis of hemophilia A.";
RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
RN [10]
RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE=89088506; PubMed=2491949;
RA White G.C. II, Shoemaker C.B.;
RT "Factor VIII gene and hemophilia A.";
RL Blood 73:1-12(1989).
RN [11]
RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE=95245332; PubMed=7728145;
RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
RL Hum. Mutat. 5:1-22(1995).
RN [12]
RP VARIANT GLN-2326.
RX MEDLINE=86235434; PubMed=3012775;
RA Gitschier J., Wood W.I., Shuman M.A., Lawn R.M.;
RT "Identification of a missense mutation in the factor VIII gene of a
RT mild hemophiliac.";
RL Science 232:1415-1416(1986).
RN [13]
RP VARIANT PRO-2135.
RX MEDLINE=88096539; PubMed=3122181;
RA Levinson B., Janco R.L., Phillips J.A. III, Gitschier J.;
RT "A novel missense mutation in the factor VIII gene identified by
RT analysis of amplified hemophilia DNA sequences.";
RL Nucleic Acids Res. 15:9797-9805(1987).
RN [14]
RP VARIANT GLN-2228.
RX MEDLINE=88191889; PubMed=2833855;
RA Youssoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
RA Kazazian H.H.;
RT "Nonsense and missense mutations in hemophilia A: estimate of the
RT relative mutation rate at CG dinucleotides.";
RL Am. J. Hum. Genet. 42:718-725(1988).
RN [15]
RP VARIANT GLY-291.
RX MEDLINE=88220354; PubMed=2835904;
RA Youssoufian H., Wong C., Aronis S., Platakoukis H., Kazazian H.H. Jr.,
RA Antonarakis S.E.;
RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
RT in exon 7 of the factor VIII gene.";
RL Am. J. Hum. Genet. 42:867-871(1988).
RN [16]
RP VARIANT CYS-1708.
RX MEDLINE=89274393; PubMed=2499363;
RA O'Brien D.P., Tuddenham E.G.;

RT "Purification and characterization of factor VIII 1,689-Cys: a
RT nonfunctional cofactor occurring in a patient with severe hemophilia
RT A.";
RL Blood 73:2117-2122(1989).
RN [17]
RP VARIANT CYS-391.
RX MEDLINE=90001543; PubMed=2506948;
RA Shima M., Ware J., Yoshioka A., Fukui H., Fulcher C.A.;
RT "An arginine to cysteine amino acid substitution at a critical
RT thrombin cleavage site in a dysfunctional factor VIII molecule.";
RL Blood 74:1612-1617(1989).
RN [18]
RP VARIANT LEU-189.
RX MEDLINE=90057680; PubMed=2510835;
RA Chan V., Chan T.K., Tong T.M., Todd D.;
RT "A novel missense mutation in exon 4 of the factor VIII:C gene
RT resulting in moderately severe hemophilia A.";
RL Blood 74:2688-2691(1989).
RN [19]
RP VARIANT LEU-2326.
RX MEDLINE=89197216; PubMed=2495245;
RA Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;
RT "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26
RT of the factor VIII gene.";
RL Hum. Genet. 81:335-338(1989).
RN [20]
RP VARIANT HIS-391.
RX MEDLINE=89264602; PubMed=2498882;
RA Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
RA Fujimaki M., Hoyer L.W.;
RT "Direct characterization of factor VIII in plasma: detection of a
RT mutation altering a thrombin cleavage site
RT (arginine-372-->histidine).";
RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
RN [21]
RP VARIANT CYS-1708.
RX MEDLINE=90105723; PubMed=2104766;
RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
RA Phillips J.A. III, Janco R.L., Hoyer L.W.;
RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
RT Cys) in the factor VIII gene of two unrelated patients with
RT cross-reacting material-positive hemophilia A.";
RL Blood 75:384-389(1990).
RN [22]
RP VARIANTS GLN-2228 AND LEU-2326.
RX MEDLINE=90123183; PubMed=2105106;
RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,
RA Mancuso G., Morfini M., de Biasi R., Baudo F., Carbonara A.;
RT "Recurrent mutations and three novel rearrangements in the factor
RT VIII gene of hemophilia A patients of Italian descent.";
RL Blood 75:662-670(1990).
RN [23]
RP VARIANT CYS-391.
RX MEDLINE=90329422; PubMed=1973901;
RA Pattinson J.K., McVey J.H., Boon M., Ajani A., Tuddenham E.G.;
RT "CRM+ haemophilia A due to a missense mutation (372-->Cys) at the
RT internal heavy chain thrombin cleavage site.";
RL Br. J. Haematol. 75:73-77(1990).
RN [24]
RP VARIANTS PHE-1699 AND CYS-1708.
RX MEDLINE=90152691; PubMed=2105906;
RA Higuchi M., Wong C., Kochhan L., Olek K., Aronis S., Kasper C.K.,
RA Kazazian H.H., Antonarakis S.E.;
RT "Characterization of mutations in the factor VIII gene by direct
RT sequencing of amplified genomic DNA.";
RL Genomics 6:65-71(1990).
RN [25]
RP VARIANTS CYS-1728 AND ASP-1941.
RX MEDLINE=90169988; PubMed=2106480;
RA Traystman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,
RA Kazazian H.H.;
RT "Use of denaturing gradient gel electrophoresis to detect point
RT mutations in the factor VIII gene.";

```
Query Match          93.4%; Score 127; DB 1; Length 2351;
Best Local Similarity 96.0%; Pred. No. 5.8e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
   || ||||| ||||| ||||| |||||
Db 503 RPLYSRRLPKGVKHLKDFPILPGEI 527

RESULT 2
FA8_MOUSE          STANDARD;          PRT; 2319 AA.
AC Q06194;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor VIII precursor (Procoagulant component).
GN F8 OR CF8 OR F8C.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RX MEDLINE=93300511; PubMed=8314577;
RA Elder B., Lakich D., Gitschier J.;
RT "Sequence of the murine factor VIII cDNA."
RL Genomics 16:374-379(1993).
CC -!- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS
CC AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE
CC ACTIVATED FORM, FACTOR XA.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: FOUND IN MOST TISSUES.
CC -!- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L05573; AAA37385.1; -.
DR PIR; A47004; A47004.
DR HSSP; P00451; ICFG.
DR MGD; MGI:88383; F8.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
KW Blood coagulation; Repeat; Plasma; Acute phase; Calcium;
KW Signal; Glycoprotein; Sulfation.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2319 COAGULATION FACTOR VIII.
FT DOMAIN 20 349 F5/8 TYPE A 1.
FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.
FT DOMAIN 207 349 PLASTOCYANIN-LIKE 2.
FT DOMAIN 399 730 F5/8 TYPE A 2.
FT DOMAIN 399 573 PLASTOCYANIN-LIKE 3.
FT DOMAIN 583 730 PLASTOCYANIN-LIKE 4.
FT DOMAIN 760 1640 B.
FT DOMAIN 1683 2008 F5/8 TYPE A 3.
FT DOMAIN 1683 1845 PLASTOCYANIN-LIKE 5.
```

```
FT DOMAIN 1855 2008 PLASTOCYANIN-LIKE 6.
FT DOMAIN 2009 2156 F5/8 TYPE C 1.
FT DOMAIN 2161 2313 F5/8 TYPE C 2.
FT SITE 391 392 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 759 760 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1678 1679 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1324 1325 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
FT SITE 1640 1641 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
FT MOD_RES 367 367 SULFATION (BY SIMILARITY).
FT MOD_RES 737 737 SULFATION (BY SIMILARITY).
FT MOD_RES 738 738 SULFATION (BY SIMILARITY).
FT MOD_RES 742 742 SULFATION (BY SIMILARITY).
FT MOD_RES 1669 1669 SULFATION (REQUIRED FOR VWF BINDING)
FT (BY SIMILARITY).
FT MOD_RES 1687 1687 SULFATION (BY SIMILARITY).
FT DISULFID 173 199 PROBABLE.
FT DISULFID 547 573 PROBABLE.
FT DISULFID 1819 1845 PROBABLE.
FT DISULFID 2008 2156 BY SIMILARITY.
FT DISULFID 2161 2313 BY SIMILARITY.
FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 601 601 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 880 880 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 958 958 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1015 1015 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1026 1026 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1076 1076 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1087 1087 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1136 1136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1161 1161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1192 1192 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1255 1255 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1268 1268 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1273 1273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1302 1302 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1316 1316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1340 1340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1797 1797 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2105 2105 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 2319 AA; 266148 MW; FD054DE051DB2A01 CRC64;
```

```
Query Match          69.1%; Score 94; DB 1; Length 2319;
Best Local Similarity 66.7%; Pred. No. 6.1e-06;
Matches 16; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 2 PNYSRRLPKGVKHLKDFPILPGEI 25
   | ::|||:|:|:| | | |||
Db 504 PLHARRLPKGIKHVKDLPIHPGEI 527
```

```
RESULT 3
FA8_PIG          STANDARD;          PRT; 2133 AA.
ID FA8_PIG
AC P12263; Q95243;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor VIII precursor (Procoagulant component).
GN F8 OR CF8.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Healey J.F., Lubin I.M., Iollar P.;
```


RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 705-1573 FROM N.A.
RX MEDLINE=86287369; PubMed=3016730;
RA Toole J.J., Pittman D.D., Orr E.C., Murtha P., Wasley L.C.,
RA Kaufman R.J.;
RT "A large region (approximately equal to 95 kDa) of human factor VIII
is dispensable for in vitro procoagulant activity.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5939-5942(1986).
RN [3]
RP SEQUENCE OF 392-759 FROM N.A.
RX MEDLINE=94179260; PubMed=7510693;
RA Lubin I.M., Healey J.F., Scandella D., Runge M.S., Lollar P.;
RT "Elimination of a major inhibitor epitope in factor VIII.";
RL J. Biol. Chem. 269:8639-8641(1994).
CC -!- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS
AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE
CC ACTIVATED FORM, FACTOR XA.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
2 PLASTOCYANIN-LIKE REPEATS.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U49517; AAB06705.1; -.
DR PIR; A25945; A25945.
DR HSSP; P00451; 1CFG.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
KW Blood coagulation; Repeat; Plasma; Acute phase; Calcium;
KW Signal; Glycoprotein; Sulfation.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2133 COAGULATION FACTOR VIII.
FT DOMAIN 20 357 F5/8 TYPE A 1.
FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.
FT DOMAIN 207 357 PLASTOCYANIN-LIKE 2.
FT DOMAIN 399 730 F5/8 TYPE A 2.
FT DOMAIN 399 573 PLASTOCYANIN-LIKE 3.
FT DOMAIN 583 730 PLASTOCYANIN-LIKE 4.
FT DOMAIN 760 1599 B.
FT DOMAIN 1495 1822 F5/8 TYPE A 3.
FT DOMAIN 1495 1659 PLASTOCYANIN-LIKE 5.
FT DOMAIN 1669 1822 PLASTOCYANIN-LIKE 6.
FT DOMAIN 1823 1970 F5/8 TYPE C 1.
FT DOMAIN 1975 2127 F5/8 TYPE C 2.
FT SITE 391 392 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 759 760 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1449 1450 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
FT SITE 1490 1491 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT MOD_RES 737 737 SULFATION (BY SIMILARITY).
FT MOD_RES 738 738 SULFATION (BY SIMILARITY).
FT MOD_RES 742 742 SULFATION (BY SIMILARITY).
FT DISULFID 173 199 PROBABLE.
FT DISULFID 547 573 PROBABLE.
FT DISULFID 1633 1659 PROBABLE.
FT DISULFID 1822 1970 BY SIMILARITY.
FT DISULFID 1975 2127 BY SIMILARITY.
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	601	601	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	929	929	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	985	985	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1025	1025	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1111	1111	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1181	1181	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1208	1208	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1245	1245	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1265	1265	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1335	1335	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1408	1408	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1611	1611	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1919	1919	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CONFLICT	713	713	N -> M (IN REF. 2).	
FT	CONFLICT	734	734	I -> T (IN REF. 2).	
FT	CONFLICT	792	792	G -> Q (IN REF. 2).	
FT	CONFLICT	1133	1133	E -> F (IN REF. 2).	
FT	CONFLICT	1191	1191	I -> L (IN REF. 2).	
FT	CONFLICT	1209	1209	R -> F (IN REF. 2).	
FT	CONFLICT	1437	1437	C -> G (IN REF. 2).	
FT	CONFLICT	1456	1456	F -> R (IN REF. 2).	
FT	CONFLICT	1539	1539	F -> R (IN REF. 2).	
FT	CONFLICT	1546	1546	Q -> N (IN REF. 2).	
SQ	SEQUENCE	2133	AA; 239304	MW; 152BBA8997F570DA	CRC64;

Query Match 55.1%; Score 75; DB 1; Length 2133;
Best Local Similarity 83.3%; Pred. No. 0.0044;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy	7	RLPKGVKHLKDFPILPGE 24
Db	509	RLKKGWKHLKDMPIPGE 526

RESULT 4
PUT4_HUMAN
ID FUT4_HUMAN STANDARD; PRT; 405 AA.
AC P22083;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alpha-(1,3)-fucosyltransferase (EC 2.4.1.-) (Galactoside 3-L-
DE fucosyltransferase) (Fucosyltransferase 4) (FUCT-IV) (ELAM-1 ligand
DE fucosyltransferase).
GN FUT4 OR ELFT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Periphereal blood leukocytes;
RX MEDLINE=91373370; PubMed=1716630;
RA Lowe J.B., Kukowska-Latallo J.F., Nair R.P., Larsen R.D., Marks R.M.,
RA Macher B.A., Kelly R.J., Ernst L.K.;
RT "Molecular cloning of a human fucosyltransferase gene that determines
RT expression of the Lewis x and VIM-2 epitopes but not ELAM-1-dependent
RT cell adhesion.";
RL J. Biol. Chem. 266:17467-17477(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91084863; PubMed=1702034;
RA Goelz S.E., Hession C., Goff D., Griffiths B., Tizard R., Newman B.,
RA Chi-Rosso G., Lobb R.;
RT "ELFT: a gene that directs the expression of an ELAM-1 ligand.";
RL Cell 63:1349-1356(1990).
RN [3]
RP SEQUENCE OF 1-400 FROM N.A.
RX MEDLINE=92042084; PubMed=1718983;
RA Kumar R., Potvin B., Muller W.A., Stanley P.;
RT "Cloning of a human alpha(1,3)-fucosyltransferase gene that encodes
RT ELFT but does not confer ELAM-1 recognition on Chinese hamster ovary
RT cell transfectants.";

RL J. Biol. Chem. 266:21777-21783(1991).
CC -!- FUNCTION: MAY CATALYSE ALPHA-1,3 GLYCOSIDIC LINKAGES INVOLVED IN
CC THE EXPRESSION OF LEWIS X/SSEA-1 AND VIM-2 ANTIGENS.
CC
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC FORM IN TRANS CISTERNAE OF GOLGI.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 10.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M65030; AAA92977.1; -.
CC EMBL; M58596; AAA63172.1; -.
CC EMBL; M58597; AAA63173.1; ALT_INIT.
CC EMBL; S65161; AAB20349.1; -.
CC PIR; A36340; A36340.
CC Genew; HGNC:4015; FUT4.
CC MIM; 104230; -.
CC InterPro; IPR001503; GT_10.
CC Pfam; PF00852; Glyco_transf_10; 1.
KW Transferase; Glycosyltransferase; Transmembrane; Glycoprotein;
KW Signal-anchor; Golgi stack.
FT DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 23 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 48 405 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 87 87 P -> R (IN REF. 2 AND 3).
FT CONFLICT 241 241 E -> D (IN REF. 3).
SQ SEQUENCE 405 AA; 45569 MW; DE72E1FDC390268D CRC64;

Query Match 35.7%; Score 48.5; DB 1; Length 405;
Best Local Similarity 50.0%; Pred. No. 7.2;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 1 RPNYSRRLPKGV-KHLKDFP 19
| | | | : | | | : | | |
Db 319 RANYERFVPRGAFIHVDFF 338

RESULT 5
FUT4_MOUSE STANDARD; PRT; 433 AA.
ID FUT4_MOUSE
AC Q11127;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alpha-(1,3)-fucosyltransferase (EC 2.4.1.-) (Galactoside 3-L-
DE fucosyltransferase) (Fucosyltransferase 4) (FUCT-IV).
GN FUT4 OR ELFT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96027607; PubMed=7559635;
RA Gersten K.M., Natsuka S., Trinchera M., Petryniak B., Kelly R.J.,
RA Hiraiwa N., Jenkins N.A., Gilbert D.J., Copeland N.G., Lowe J.B.;
RT "Molecular cloning, expression, chromosomal assignment, and tissue-
RT specific expression of a murine alpha-(1,3)-fucosyltransferase locus
RT corresponding to the human ELAM-1 ligand fucosyl transferase.";
RL J. Biol. Chem. 270:25047-25056(1995).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC STRAIN=129/Sv; TISSUE=Liver;
RX MEDLINE=97037075; PubMed=8882722;

RA Ozawa M., Muramatsu T.;
RT "Molecular cloning and expression of a mouse alpha-1,3
RT fucosyltransferase gene that shows homology with the human alpha-1,3
RT fucosyltransferase IV gene."
RL J. Biochem. 119:302-308(1996).
CC -!- FUNCTION: MAY CATALYSE ALPHA-1,3 GLYCOSIDIC LINKAGES INVOLVED IN
CC THE EXPRESSION OF LEWIS X/SSEA-1 AND VIM-2 ANTIGENS.
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC FORM IN TRANS CISTERNAE OF GOLGI.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN STOMACH AND COLON.
CC IT ALSO EXPRESSED IN THE LUNG, TESTIS, UTERUS, SMALL INTESTINE
CC AND TO A LESSER EXTENT IN SPLEEN, AND OVARY. PRESENT IN TRACE
CC AMOUNTS IN BRAIN, THYMUS, HEART, SMOOTH MUSCLE, KIDNEY AND BONE
CC MARROW. NOT FOUND IN LIVER, SALIVARY GLAND AND PANCREAS.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 10.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U33457; AAC52269.1; -.
CC EMBL; D63380; BAA09697.1; -.
CC EMBL; D63379; BAA09696.1; -.
CC MGD; MGI:95594; Fut4.
CC InterPro; IPR001503; GT_10.
CC Pfam; PF00852; Glyco_transf_10; 1.
KW Transferase; Glycosyltransferase; Transmembrane; Glycoprotein;
KW Signal-anchor; Golgi stack; Alternative splicing.
FT DOMAIN 1 52 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 53 74 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 75 433 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 33 MISSING (IN SHORT ISOFORM).
FT CONFLICT 252 252 Q -> P (IN REF. 2).
FT CONFLICT 257 257 R -> Q (IN REF. 2).
FT CONFLICT 260 260 V -> E (IN REF. 2).
FT CONFLICT 273 273 R -> Q (IN REF. 2).
SQ SEQUENCE 433 AA; 49481 MW; 2401822F02B5D021 CRC64;

Query Match 35.7%; Score 48.5; DB 1; Length 433;
Best Local Similarity 50.0%; Pred. No. 7.8;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 1 RPNYSRRLPKGV-KHLKDFP 19
| | | | : | | | : | | |
Db 347 RANYERFVPRGAFIHVDFF 366

RESULT 6
DPO2_RAT STANDARD; PRT; 495 AA.
ID DPO2_RAT
AC O89043;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B)
DE (DNA polymerase subunit II) (Fragment).
GN POLA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -|- FUNCTION: INTERCONVERSION OF SERINE AND GLYCINE.
CC -|- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + glycine +
CC H(2)O = tetrahydrofolate + L-serine.
CC -|- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -|- PATHWAY: KEY ENZYME IN THE BIOSYNTHESIS OF PURINES, LIPIDS,
CC HORMONES AND OTHER COMPONENTS.
CC -|- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -|- SIMILARITY: BELONGS TO THE SHMT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).

DR EMBL; Z98601; CAB11269.1; -.
DR HSSP; P07511; 1CJ0.
DR InterPro; IPR001085; Gly_HyMettransf.
DR Pfam; PF00464; SHMT; 1.
DR PROSITE; PS00096; SHMT; 1.
KW Transferase; Pyridoxal phosphate; One-carbon metabolism.
FT BINDING 243 243 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT DOMAIN 238 241 POLY-THR.
SQ SEQUENCE 467 AA; 51861 MW; 0C21D7EF010C3725 CRC64;

Query Match 33.8%; Score 46; DB 1; Length 467;
Best Local Similarity 45.5%; Pred. No. 20;
Matches 10; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 3 NYSRRLPKGVKHLKDFPILPGE 24
I : ||| |||| |
Db 419 NLQKELPKDANKLKDFKAKLGE 440

RESULT 9
UBIG_XYLFA
ID UBIG_XYLFA STANDARD; PRT; 246 AA.
AC Q9PAM5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-demethylubiquinone-9 3-methyltransferase (EC 2.1.1.64) (3,4-
DE dihydroxy-5-hexaprenylbenzoate methyltransferase) (DHBB
DE methyltransferase).
GN UBIG OR XF2471.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,

RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -|- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + 3-
CC demethylubiquinone-9 - S-adenosyl-L-homocysteine + ubiquinone-9.
CC -|- PATHWAY: Ubiquinone biosynthesis.
CC -|- SIMILARITY: BELONGS TO THE UBIG/COQ3 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).

DR EMBL; AE004055; AAF85269.1; -.
DR InterPro; IPR001601; Methyltransf.
DR InterPro; IPR000051; SAM_bind.
KW Ubiquinone biosynthesis; Transferase; Methyltransferase;
KW Complete proteome.
SQ SEQUENCE 246 AA; 27121 MW; 5212107D63633D3F CRC64;

Query Match 33.5%; Score 45.5; DB 1; Length 246;
Best Local Similarity 55.0%; Pred. No. 12;
Matches 11; Conservative 1; Mismatches 7; Indels 1; Gaps 1;

QY 6 RRLPKGVKHLKDFPILPGEI 25
I : |||| ||| | | |
Db 178 RLLPKGTHQYKDF-IRPAEL 196

RESULT 10
MHPA_ECOLI
ID MHPA_ECOLI STANDARD; PRT; 554 AA.
AC P77397; P71203; P77047;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-(3-hydroxy-phenyl)propionate hydroxylase (EC 1.14.13.-).
GN MHPA OR B0347.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Kawamukai M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / CS520;
RA Ferrandez A., Garcia J.L., Diaz E.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MGL1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

```
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- COFACTOR: FAD (BY SIMILARITY).
CC -!- PATHWAY: 3-hydroxyphenylpropionate degradation.
CC -!- SIMILARITY: BELONGS TO THE PHEA/TPDB FAMILY OF FAD MONOOXYGENASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D86239; BAAL3052.1; -.
CC EMBL; Y09555; CAA70747.1; -.
CC DR EMBL; AE000142; AAC73450.1; -.
CC DR EMBL; U73857; AAB18071.1; -.
CC DR EcoGene; EG20273; mhpA.
CC DR InterPro; IPR000733; Flav_monooxygenase.
CC DR InterPro; IPR002938; MoxY_FAD_binding.
CC DR InterPro; IPR000205; NAD_binding.
CC DR InterPro; IPR003042; Rng_mnnoxygenase.
CC DR Pfam; PF01360; Monooxygenase; 1.
CC DR Pfam; PF01494; FAD_binding_3; 1.
CC DR PRINTS; PR00420; RNCMNOXGNASE.
CC DR Aromatic hydrocarbons catabolism; Oxidoreductase; Flavoprotein; FAD;
KW Complete proteome.
FT NP_BIND 17 46 FAD (POTENTIAL).
FT NP_BIND 285 295 FAD (POTENTIAL).
FT CONFLICT 272 272 Q -> H (IN REF. 1).
FT CONFLICT 360 360 L -> P (IN REF. 1).
FT CONFLICT 502 502 W -> G (IN REF. 1).
SQ SEQUENCE 554 AA; 62185 MW; 1D56CB799E9F8A8E CRC64;

Query Match 33.5%; Score 45.5; DB 1; Length 554;
Best Local Similarity 41.7%; Pred. No. 30;
Matches 10; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

QY 1 RPNYSRRLPKGVKHLKDFPILPGE 24
|| | || | : | : |||
DB 220 RPYVSAALPHAVERF-EFVMVMPGE 242

RESULT 11
MURA_SYNY3
ID MURA_SYNY3 STANDARD; PRT; 438 AA.
AC Q55673;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC 2.5.1.7)
DE (Enolpyruvate transferase) (UDP-N-acetylglucosamine enolpyruvyl
DE transferase) (EPT).
GN MURA OR MUR2 OR SLR0017.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiyura M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
```

```
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -!- FUNCTION: CELL WALL FORMATION. ADDS ENOLPYRUVYL TO UDP-N-
CC ACETYLGLUCOSAMINE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + UDP-N-acetyl-D-
CC glucosamine = phosphate + UDP-N-acetyl-3-O-(1-carboxyvinyl)-D-
CC glucosamine.
CC -!- PATHWAY: Peptidoglycan biosynthesis; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY. MURA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D64000; BAA10199.1; -.
CC DR HSSP; P33038; IDLG.
CC DR InterPro; IPR001986; EPSP_synthase.
CC DR Pfam; PF00275; EPSP_synthase; 1.
CC DR ProDom; PD001867; EPSP_synthase; 1.
CC DR TIGRFAMS; TIGR01072; mura; 1.
CC KW Peptidoglycan synthesis; Cell wall; Cell division; Transferase;
CC Complete proteome.
CC FT ACT_SITE 129 129 BINDS PEP (BY SIMILARITY).
CC SQ SEQUENCE 438 AA; 46604 MW; C310AB287F2F0F55 CRC64;

Query Match 33.1%; Score 45; DB 1; Length 438;
Best Local Similarity 47.8%; Pred. No. 27;
Matches 11; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

QY 5 SRRLPKGVK--HLKDFPILPGEI 25
|| : || : | ||||:| |
DB 228 SRIINGVEKLHSTDFPIIPDRI 250

RESULT 12
ITA4_XENLA
ID ITA4_XENLA STANDARD; PRT; 1032 AA.
AC Q91687; Q06273;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-4 precursor (Integrin alpha-IV) (VLA-4).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96379747; PubMed=8787760;
RA Ramos J.W., Whittaker C.A., Desimone D.W.;
RT "Integrin-dependent adhesive activity is spatially controlled by
RT inductive signals at gastrulation.";
RL Development 122:2873-2883(1996).
RN [2]
RP SEQUENCE OF 308-379 FROM N.A.
RX MEDLINE=94008528; PubMed=8404528;
RA Whittaker C.A., Desimone D.W.;
RT "Integrin alpha subunit mRNAs are differentially expressed in early
RT Xenopus embryos.";
RL Development 117:1239-1249(1993).
CC -!- FUNCTION: FIBRONECTIN AND V-CAM ADHESION RECEPTOR.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -!- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```



```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U54497; AAA98673.1; -.
DR EMBL; L10188; AAA16248.1; -.
DR HSSP; P11215; IABX.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF00357; integrin_A; 1.
DR Pfam; PF01839; FG-GAP; 5.
DR PRINTS; PR01185; INTEGRINA.
DR SMART; SM00191; Int_alpha; 5.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Repeat.
FT SIGNAL 1 34 POTENTIAL.
FT CHAIN 35 1032 INTEGRIN ALPHA-4.
FT DOMAIN 35 974 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 975 998 POTENTIAL.
FT DOMAIN 999 1032 CYTOPLASMIC (POTENTIAL).
FT REPEAT 50 112 FG-GAP 1.
FT REPEAT ? ? FG-GAP 2.
FT REPEAT 197 247 FG-GAP 3.
FT REPEAT 248 301 FG-GAP 4.
FT REPEAT 303 361 FG-GAP 5.
FT REPEAT 364 423 FG-GAP 6.
FT REPEAT 426 474 FG-GAP 7.
FT CA_BIND 314 322 POTENTIAL.
FT CA_BIND 376 384 POTENTIAL.
FT CA_BIND 438 446 POTENTIAL.
FT SITE 590 591 CLEAVAGE.
FT SITE 1001 1005 GFFKR MOTIF.
FT DISULFID 91 101 BY SIMILARITY.
FT DISULFID 144 165 BY SIMILARITY.
FT DISULFID 183 198 BY SIMILARITY.
FT DISULFID 485 494 BY SIMILARITY.
FT DISULFID 500 556 BY SIMILARITY.
FT DISULFID 622 627 BY SIMILARITY.
FT DISULFID 698 712 BY SIMILARITY.
FT DISULFID 853 889 BY SIMILARITY.
FT DISULFID 896 901 BY SIMILARITY.
FT CARBOHYD 81 81 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 229 229 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 479 479 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 496 496 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 517 517 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 537 537 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 626 626 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 660 660 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 746 746 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1032 AA; 115215 MW; 6486797D83AAE69E CRC64;

Query Match 33.1%; Score 45; DB 1; Length 1032;
Best Local Similarity 52.6%; Pred. No. 71;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 7 RLPKGVKHLKDFPILPGEI 25
|||||: :| :| :|
Db 678 RLPKGLYFVKVFDLLEKEI 696

RESULT 13
CA64_HUMAN STANDARD; PRT; 1678 AA.
AC Q14031;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
```

```
DE Collagen alpha 6(IV) chain precursor.
GN COL4A6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye, and Kidney;
RX MEDLINE=94171779; PubMed=8125972;
RA Oohashi T., Sugimoto M., Mattei M.-G., Ninomiya Y.;
RT "Identification of a new collagen IV chain, alpha 6(IV), by cDNA
RT isolation and assignment of the gene to chromosome Xq22, which is the
RT same locus for COL4A5.";
RL J. Biol. Chem. 269:7520-7526(1994).
CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
CC NIDOGEN.
CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D21337; BAA04809.1; -.
DR Genew; HGNC:2208; COL4A6.
DR MIM; 303631; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001442; ProcollagnC4.
DR Pfam; PF01391; Collagen; 23.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD000007; Collagen; 1.
DR ProDom; PD003923; ProcollagnC4; 2.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 1678 COLLAGEN ALPHA 6(IV) CHAIN.
FT DOMAIN 16 45 7S DOMAIN.
FT DOMAIN 46 1450 TRIPLE-HELICAL REGION.
FT DOMAIN 1451 1678 NONHELICAL REGION (NC1).
FT SITE 514 516 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 559 561 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 985 987 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 1469 1558 OR 1555 (BY SIMILARITY).
FT DISULFID 1502 1555 OR 1558 (BY SIMILARITY).
FT DISULFID 1514 1520 BY SIMILARITY.
FT DISULFID 1577 1674 OR 1671 (BY SIMILARITY).
FT DISULFID 1611 1671 OR 1674 (BY SIMILARITY).
FT DISULFID 1623 1630 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1678 AA; 162461 MW; 65E4EA282D3D37BD CRC64;
```

Query Match		33.1%	Score 45;	DB 1;	Length 1678;																	
Best Local Similarity		56.2%	Pred. No. 1.2e+02;																			
Matches	9;	Conservative	2;	Mismatches	5;	Indels	0;															
				Gaps		0;																
QY	9	PKG VKHLKDFPILPGE	24																			
		: :																				
Db	715	PRGEKGLPGFGLPGK	730																			
RESULT 14																						
CA44_HUMAN			STANDARD;	PRT;	1690	AA.																
ID	CA44_HUMAN																					
AC	P53420;																					
DT	01-OCT-1996	(Rel. 34, Created)																				
DT	01-OCT-1996	(Rel. 34, Last sequence update)																				
DT	15-JUN-2002	(Rel. 41, Last annotation update)																				
DE	Collagen alpha 4(IV) chain precursor.																					
GN	COL4A4.																					
OS	Homo sapiens (Human).																					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;																					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.																					
OX	NCBI_TaxID=9606;																					
RN	[1]																					
RP	SEQUENCE FROM N.A.																					
RC	TISSUE=Kidney;																					
RX	MEDLINE=95014445; PubMed=7523402;																					
RA	Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Reeders S.T.;																					
RT	"Complete primary structure of the human type IV collagen alpha 4(IV)																					
RT	chain. Comparison with structure and expression of the other alpha																					
RT	(IV) chains.";																					
RL	J. Biol. Chem. 269:26172-26177(1994).																					
RN	[2]																					
RP	SEQUENCE OF 1-23 FROM N.A.																					
RX	MEDLINE=98196854; PubMed=9537506;																					
RA	Momota R., Sugimoto M., Oohashi T., Kigasawa K., Yoshioka H.,																					
RA	Ninomiya Y.;																					
RT	"Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and																					
RT	alpha4(IV) collagen chains are arranged head-to-head on chromosome																					
RT	2q36.";																					
RL	FEBS Lett. 424:11-16(1998).																					
RN	[3]																					
RP	SEQUENCE OF 1219-1690 FROM N.A.																					
RC	TISSUE=Eye;																					
RX	MEDLINE=93374047; PubMed=8365481;																					
RA	Sugimoto M., Oohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;																					
RT	"cDNA isolation and partial gene structure of the human alpha 4(IV)																					
RT	collagen chain.";																					
RL	FEBS Lett. 330:122-128(1993).																					
RN	[4]																					
RP	SEQUENCE OF 1407-1507 FROM N.A.																					
RX	MEDLINE=93054733; PubMed=1429714;																					
RA	Kamagata Y., Mattei M.-G., Ninomiya Y.;																					
RT	"Isolation and sequencing of cDNAs and genomic DNAs encoding the																					
RT	alpha 4 chain of basement membrane collagen type IV and assignment of																					
RT	the gene to the distal long arm of human chromosome 2.";																					
RL	J. Biol. Chem. 267:23753-23758(1992).																					
RN	[5]																					
RP	REVIEW ON VARIANTS.																					
RX	MEDLINE=97338662; PubMed=9195222;																					
RA	Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;																					
RT	"The clinical spectrum of type IV collagen mutations.";																					
RL	Hum. Mutat. 9:477-499(1997).																					
RN	[6]																					
RP	VARIANT AS SER-1201.																					
RX	MEDLINE=95078927; PubMed=7987396;																					
RA	Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,																					
RA	Pirson Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,																					
RA	Smeets H.J.M., Reeders S.T.;																					
RT	"Identification of mutations in the alpha 3(IV) and alpha 4(IV)																					
RT	collagen genes in autosomal recessive Alport syndrome.";																					
RL	Nat. Genet. 8:77-82(1994).																					
RN	[7]																					
RP	VARIANT BFH GLU-897.																					
RX	MEDLINE=96379660; PubMed=8787673;																					
RA	Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,																					
RA	Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;																					
RT	"Benign familial hematuria due to mutation of the type IV collagen																					
RT	alpha4 gene.";																					
RL	J. Clin. Invest. 98:1114-1118(1996).																					
RN	[8]																					
RP	VARIANTS AS, AND VARIANTS.																					
RX	MEDLINE=99011253; PubMed=9792860;																					
RA	Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,																					
RA	Cochat P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;																					
RT	"Determination of the genomic structure of the COL4A4 gene and of																					
RT	novel mutations causing autosomal recessive Alport syndrome.";																					
RL	Am. J. Hum. Genet. 63:1329-1340(1998).																					
CC	-!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF																					
CC	GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'																					
CC	MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/																					
CC	NIDOGEN.																					
CC	-!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-																					
CC	ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE																					
CC	WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.																					
CC	-!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).																					
CC	-!- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE																					
CC	COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,																					
CC	COCHLEA, LUNG AND BRAIN.																					
CC	-!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS																					
CC	DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE																					
CC	G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY																					
CC	CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL																					
CC	TRIPLE-HELICAL 7S DOMAIN.																					
CC	-!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING																					
CC	UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.																					
CC	-!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH																					
CC	ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF																					
CC	THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE																					
CC	IV COLLAGENS.																					
CC	-!- DISEASE: DEFECTS IN COL4A4 ARE A CAUSE OF TYPE II AUTOSOMAL																					
CC	RECESSIVE FORM OF ALPORT SYNDROME (AS), AN HEREDITARY																					
CC	GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,																					
CC	HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN																					
CC	MALES AND FEMALES.																					
CC	-!- DISEASE: DEFECTS IN COL4A4 ARE A CAUSE OF FAMILIAL BENIGN																					
CC	HEMATURIA (FBH) OR THIN BASEMENT MEMBRANE DISEASE. FBH IS																					
CC	CHARACTERIZED BY PERSISTENT HEMATURIA, AN ELECTRON MICROSCOPICALLY																					
CC	DETECTABLE THIN GLOMERULAR BASEMENT MEMBRANE (GBM) AND AN																					
CC	AUTOSOMAL DOMINANT MODE OF INHERITANCE. RENAL FUNCTION REMAINS																					
CC	NORMAL. IN CHILDREN DIFFERENTIATION BETWEEN FBH AND AS CAN BE																					
CC	DIFFICULT, BECAUSE BOTH DISORDERS ARE MANIFESTED BY PERSISTENT																					
CC	HEMATURIA AND THIN GBM AT THAT AGE.																					
CC	-!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.																					
CC	-----																					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration																					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -																					
CC	the European Bioinformatics Institute. There are no restrictions on its																					
CC	use by non-profit institutions as long as its content is in no way																					
CC	modified and this statement is not removed. Usage by and for commercial																					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/																					
CC	or send an email to license@isb-sib.ch).																					
CC	-----																					
DR	EMBL; X81053; CAA56943.1; --.																					
DR	EMBL; AB008496; BAA25065.1; --.																					
DR	EMBL; D17391; BAA04214.1; --.																					
DR	Genew; HGNC:2206; COL4A4.																					
DR	MIM; 120131; --.																					
DR	MIM; 141200; --.																					
DR	MIM; 203780; --.																					
DR	InterPro; IPR000087; Collagen.																					
DR	InterPro; IPR001442; ProcollagnC4.																					
DR	Pfam; PF01391; Collagen; 21.																					
DR	Pfam; PF01413; C4; 2.																					
DR	ProDom; PD000007; Collagen; 1.																					
DR	ProDom; PD003923; ProcollagnC4; 2.																					
DR	SMART; SM00111; C4; 2.																					

KW Extracellular matrix; Connective tissue; Basement membrane; Repeat;
KW Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;
KW Polymorphism; Alport syndrome.
FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 1690 COLLAGEN ALPHA 4(IV) CHAIN.
FT DOMAIN 39 64 7S DOMAIN.
FT DOMAIN 65 1459 TRIPLE-HELICAL REGION.
FT DOMAIN 1460 1690 NONHELICAL REGION (NC1).
FT SITE 94 96 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 145 147 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 189 191 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 310 312 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 724 726 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 785 787 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 989 991 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1206 1206 CLEAVAGE (BY COLLAGENASE)
FT SITE 1212 1214 (BY SIMILARITY).
FT SITE 1212 1214 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 1480 1569 OR 1566 (BY SIMILARITY).
FT DISULFID 1513 1566 OR 1569 (BY SIMILARITY).
FT DISULFID 1525 1531 BY SIMILARITY.
FT DISULFID 1588 1686 OR 1683 (BY SIMILARITY).
FT DISULFID 1622 1683 OR 1686 (BY SIMILARITY).
FT DISULFID 1634 1641 BY SIMILARITY.
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 441 446 MISSING (IN AS).
FT VARIANT 545 545 /FTid=VAR_008148.
FT VARIANT 545 545 G -> A.
FT VARIANT 570 570 /FTid=VAR_008149.
FT VARIANT 897 897 E -> Q.
FT VARIANT 897 897 G -> E (IN FBH).
FT VARIANT 931 931 /FTid=VAR_001912.
FT VARIANT 1004 1004 A -> T.
FT VARIANT 1004 1004 /FTid=VAR_008151.
FT VARIANT 1030 1030 L -> P.
FT VARIANT 1201 1201 /FTid=VAR_008152.
FT VARIANT 1201 1201 G -> V (IN AS).
FT VARIANT 1402 1402 /FTid=VAR_008153.
FT VARIANT 1402 1402 G -> S (IN AS).
FT VARIANT 1572 1572 /FTid=VAR_001913.
FT VARIANT 1572 1572 P -> S.
FT VARIANT 1659 1660 /FTid=VAR_008154.
FT VARIANT 1690 1690 P -> L (IN AS).
FT CONFLICT 1659 1660 /FTid=VAR_008155.
FT SEQUENCE 1690 AA; 164095 MW; E1E72F283A72BAAE CRC64; LQ -> FE (IN REF. 3).

Query Match 33.1%; Score 45; DB 1; Length 1690;
Best Local Similarity 56.2%; Pred. No. 1.2e+02;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 9 PKGVKHLKDFPILPGE 24
| | | : | | | | |
Db 904 PKGPRGLPGFPGFGE 919

RESULT 15
YCF2_PINTH
ID YCF2_PINTH STANDARD; PRT: 2054 AA.
AC P41653;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 244.6 kDa protein ycf2 (ORF 2054).
GN YCF2.
OS Pinus thunbergii (Green pine) (Japanese black pine).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3350;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=95024047; PubMed=7937893;
RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA Sugiura M.;
RT "Loss of all ndh genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC -!- FUNCTION: NOT YET KNOWN.
CC -!- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D17510; BAA04460.1; -.
DR InterPro: IPR003959; AAA_ATPase_centra.
DR Pfam: PF00004; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2054 AA; 244604 MW; 6F5E92D078E33A9A CRC64;

Query Match 33.1%; Score 45; DB 1; Length 2054;
Best Local Similarity 61.5%; Pred. No. 1.6e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RPNYSRRRLPKGVK 13
| | | | | | | | : |
Db 786 RPNYRNRLPSDMK 798

Search completed: January 23, 2003, 09:01:12
Job time : 14 secs


```

OX NCBI_TaxID=285;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TA441;
RX MEDLINE=20005595; PubMed=10537203;
RA Arai H., Yamamoto T., Ohishi T., Shimizu T., Nakata T., Kudo T.;
RT "Genetic organization and characterization of the 3-(3-
RT hydroxyphenyl)propionic acid degradation pathway of Comamonas
RL testosteroni TA441."
RL Microbiology 145:2813-2820(1999).
CC -1- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AB024335; BAA82878.1; -.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR000733; Flav_monooxygenase.
DR InterPro; IPR002114; HPr_Serp_site.
DR InterPro; IPR002938; MoxF_FAD_binding.
DR InterPro; IPR000205; NAD_binding.
DR InterPro; IPR00103; Pyridine_redox.2.
DR InterPro; IPR003042; Rng_monooxygenase.
DR InterPro; IPR000594; Thif_domain.
DR Pfam; PF01494; FAD_binding_3; 1.
DR Pfam; PF01360; Monooxygenase; 1.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00469; PNRDRTASEII.
DR PRINTS; PR00420; RINGMONOXGNASE.
DR PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
KW FAD; Flavoprotein; Oxidoreductase.
SQ SEQUENCE 589 AA; 65297 MW; 7CACEF905825CB744 CRC64;

Query Match 37.1%; Score 50.5; DB 2; Length 589;
Best Local Similarity 41.7%; Pred. No. 29;
Matches 10; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

QY 1 RPNYSRRLPKGVKHLKDFPILPGE 24
DB 218 RPYVSAALPHGIRRF-EFMVMPGE 240

RESULT 6
Q9VE79 PRELIMINARY; PRT; 935 AA.
AC Q9VE79;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG14309 protein.
GN CG14309.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

```

```

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003721; AAF55548.1; -.
DR FlyBase; FBgn0038611; CG14309.
SQ SEQUENCE 935 AA; 106883 MW; 10E46B721CE5DB3F CRC64;

Query Match 37.1%; Score 50.5; DB 5; Length 935;
Best Local Similarity 45.8%; Pred. No. 48;
Matches 11; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

QY 2 PNYSRRLPKGVKHLKDFPILPGEI 25
DB 575 PTYT-RLPEGDVHLKHIETVDGEV 597

RESULT 7
Q8SAW1 PRELIMINARY; PRT; 388 AA.
AC Q8SAW1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative wall-associated protein kinase.
GN OSJNBA0051J07.1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA Sasaki C., Henry D., Oates R., Simmons J.;
RT "Rice Genomic Sequence."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC098566; AAL77114.1; -.
KW Kinase.
SQ SEQUENCE 388 AA; 42179 MW; 0F4F47E6F087FF91 CRC64;

Query Match 36.8%; Score 50; DB 10; Length 388;
Best Local Similarity 47.6%; Pred. No. 22;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 3 NYSRRLPKGVKHLKDFPILPG 23
DB 240 NFSKKYPKGVPLVIDFAIRDG 260

RESULT 8
Q96A81 PRELIMINARY; PRT; 579 AA.
AC Q96A81;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

```

```

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CDNA FLJ31952 fis, clone NT2RP7007221, weakly similar to Rattus
DE norvegicus schlafen-4 (SLFN-4) mRNA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK056514; BAB71201.1; -;
SQ SEQUENCE 579 AA; 65949 MW; 195AEFB2F4235C84 CRC64;

Query Match 36.8%; Score 50; DB 4; Length 579;
Best Local Similarity 40.0%; Pred. No. 34;
Matches 12; Conservative 5; Mismatches 5; Indels 8; Gaps 2;

QY 1 RPNYSRRLPKGVKHLKD----FPILPGEI 25
DQ 61 RPNYSK--KGLEHKKADLQOHLFPVPPGHL 87

RESULT 9
Q855J1 PRELIMINARY; PRT; 706 AA.
ID Q855J1
AC Q855J1
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative wall-associated protein kinase.
GN OSJNBA0011A24.33.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA McCombie W.R., de la Bastide M., Spiegel L., Preston R., Ferraro K.,
RA Kuit K., Nascimento L., Zutavern T., Ballja V., Bell M., Baker J.,
RA Santos L., Miller B., Katzenberger F., Muller S., King L., Yang C.,
RA Dike S., O'Shaughnessy A., Palmer L., Dedhia N.;
RT "Genomic sequence for Oryza sativa, Nipponbare strain, clone
RT OSJNBA0011A24, from chromosome 10, complete sequence.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC113336; AAM01181.1; -;
KW Kinase.
SQ SEQUENCE 706 AA; 77555 MW; 91E6F07E9B90BFEB CRC64;

Query Match 36.8%; Score 50; DB 10; Length 706;
Best Local Similarity 47.6%; Pred. No. 42;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 3 NYSRRLPKGVKHLKDFPILPG 23
DQ 240 NESKKYPKGVPLVDFAIRDG 260

RESULT 10
Q9PBA7 PRELIMINARY; PRT; 958 AA.
ID Q9PBA7
AC Q9PBA7
DT 01-OCT-2000 (TReMBLrel. 15, Created)

```

```

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein Xf2237.
GN Xf2237.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.B., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Queggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL; AE004036; AAF85036.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 958 AA; 103030 MW; 4EF720E00D333EDB CRC64;

Query Match 36.8%; Score 50; DB 16; Length 958;
Best Local Similarity 47.1%; Pred. No. 59;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 6 RRLPKGVKHLKDFPILP 22
DQ 124 RRLPKGVKHLKDFPILP 140

RESULT 11
Q9VP88 PRELIMINARY; PRT; 2444 AA.
ID Q9VP88
AC Q9VP88;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE Putative PAP protein (CG9936 protein).
GN PAP OR CG9936.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Calle R.F.,

```


SQ SEQUENCE 209 AA; 24508 MW; 1B99EA1727E634D3 CRC64;
 Query Match 36.0%; Score 49; DB 11; Length 209;
 Best Local Similarity 52.9%; Pred. No. 16;
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 3 NYSRRRLPKGVKHLKDFP 19
 Db 153 NFSKSIYKYVNVWKKDFP 169
 RESULT 14
 O58994 PRELIMINARY; PRT; 412 AA.
 AC O58994;
 DT 01-AUG-1998 (TReMBLrel. 07, Created)
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE Hypothetical protein PH1259.
 GN PH1259.
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=53953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OT3;
 RX MEDLINE=98344137; PubMed=9679194;
 RA Kawaiyabasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
 Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
 Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
 Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushida N., Oguchi A.,
 Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
 Masuchi Y., Shizuya H., Kikuchi H.,
 "Complete sequence and gene organization of the genome of a hyper-
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
 RL DNA Res. 5:55-76(1998).
 DR EMBL; AP000005; BAA30361.1;
 DR InterPro; IPR000051; SAM_bind.
 DR InterPro; IPR001566; TrmA.
 DR PROSITE; PS01231; TRMA_2; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 412 AA; 47094 MW; FFB2CDD11EF5B331 CRC64;
 Query Match 36.0%; Score 49; DB 17; Length 412;
 Best Local Similarity 58.8%; Pred. No. 33;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 2 PNYSRRLPKGVKHLKDF 18
 Db 140 PFSERTPKYLKALKDF 156
 RESULT 15
 O920V9 PRELIMINARY; PRT; 386 AA.
 AC O920V9;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE Alpha 1,3-fucosyltransferase (Fragment).
 GN FUT4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BLG2/MSF;
 RA Liu Y., Kitano T., Koide T., Shiroishi T., Moriaki K., Saitou N.,
 "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
 Five Mus musculus subspecies.";
 RT Five Mus musculus subspecies.";
 NT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RL

DR EMBL; AB039126; BAB68650.1; -.
 DR InterPro; IPR001503; GT_10.
 DR Pfam: PF00852; Glyco_transf_10; 1.
 KW Glycosyltransferase; Transferase.
 FT NON_TER 1 386
 FT NON_TER 386
 SQ SEQUENCE 386 AA; 43812 MW; 6E057C1A29F99A70 CRC64;
 Query Match 35.7%; Score 48.5; DB 11; Length 386;
 Best Local Similarity 50.0%; Pred. No. 37;
 Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;
 QY 1 RPNYSRRLPKGV-KHLKDFP 19
 Db 307 RANERFVPRGAFIHVDFFP 326
 Search completed: January 23, 2003, 09:03:27
 Job time : 33 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 23, 2003, 09:01:58 ; Search time 14 Seconds
(without alignments)
52.541 Million cell updates/sec

Title: SEQ2-ASNAT3
Perfect score: 136
Sequence: 1 RPNYSRRLLPKGVKHLKDFPILPGEI 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	93.4	1438	4 US-09-209-916-1	Sequence 1, Appli
2	127	93.4	1471	1 US-08-683-839B-3	Sequence 3, Appli
3	127	93.4	1661	2 US-08-882-083-2	Sequence 2, Appli
4	127	93.4	1661	2 US-08-558-107-2	Sequence 2, Appli
5	127	93.4	1661	4 US-09-243-539-2	Sequence 2, Appli
6	127	93.4	2332	1 US-07-864-004B-4	Sequence 4, Appli
7	127	93.4	2332	1 US-08-251-937A-4	Sequence 2, Appli
8	127	93.4	2332	1 US-08-212-133A-2	Sequence 2, Appli
9	127	93.4	2332	1 US-08-276-594A-2	Sequence 2, Appli
10	127	93.4	2332	1 US-08-474-503-2	Sequence 2, Appli
11	127	93.4	2332	2 US-08-670-707A-2	Sequence 2, Appli
12	127	93.4	2332	4 US-09-037-601-2	Sequence 2, Appli
13	127	93.4	2332	4 US-09-324-867-3	Sequence 3, Appli
14	127	93.4	2332	4 US-09-315-179-2	Sequence 2, Appli
15	127	93.4	2332	4 US-09-523-656-2	Sequence 2, Appli
16	127	93.4	2332	5 PCT-US93-03275-4	Sequence 4, Appli
17	127	93.4	2332	5 PCT-US94-13200-2	Sequence 2, Appli
18	127	93.4	2351	1 US-08-121-202-2	Sequence 2, Appli
19	127	93.4	2351	1 US-08-366-851A-2	Sequence 2, Appli
20	127	93.4	2351	6 5171844-2	Patent No. 5171844
21	127	93.4	2351	6 5422260-1	Patent No. 5422260
22	101	74.3	2343	4 US-09-324-867-2	Sequence 2, Appli
23	94	69.1	2304	4 US-09-324-867-4	Sequence 4, Appli
24	94	69.1	2319	1 US-08-212-133A-8	Sequence 8, Appli
25	94	69.1	2319	1 US-08-474-503-6	Sequence 6, Appli
26	94	69.1	2319	2 US-08-670-707A-6	Sequence 6, Appli
27	94	69.1	2319	4 US-09-037-601-6	Sequence 6, Appli

28	94	69.1	2319	4 US-09-315-179-6	Sequence 6, Appli
29	94	69.1	2319	4 US-09-523-656-28	Sequence 28, Appli
30	94	69.1	2319	5 PCT-US94-13200-6	Sequence 6, Appli
31	75	55.1	367	1 US-07-864-004B-2	Sequence 2, Appli
32	75	55.1	367	1 US-08-251-937A-2	Sequence 2, Appli
33	75	55.1	367	5 PCT-US93-03275-2	Sequence 2, Appli
34	75	55.1	368	1 US-08-212-133A-6	Sequence 6, Appli
35	75	55.1	368	1 US-08-474-503-4	Sequence 4, Appli
36	75	55.1	368	2 US-08-670-707A-4	Sequence 4, Appli
37	75	55.1	368	4 US-09-037-601-4	Sequence 4, Appli
38	75	55.1	368	4 US-09-315-179-4	Sequence 4, Appli
39	75	55.1	368	4 US-09-523-656-4	Sequence 4, Appli
40	75	55.1	368	5 PCT-US94-13200-4	Sequence 4, Appli
41	75	55.1	541	1 US-08-121-202-4	Sequence 4, Appli
42	75	55.1	1443	2 US-08-670-707A-39	Sequence 39, Appli
43	75	55.1	1443	4 US-09-037-601-39	Sequence 39, Appli
44	75	55.1	1443	4 US-09-315-179-39	Sequence 39, Appli
45	75	55.1	1467	4 US-09-523-656-38	Sequence 38, Appli

ALIGNMENTS

RESULT 1
US-09-209-916-1
; Sequence 1, Application US/09209916
; Patent No. 6358703
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255
; CURRENT APPLICATION NUMBER: US/09/209,916
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-09-209-916-1

Query Match 93.4%; Score 127; DB 4; Length 1438;
Best Local Similarity 96.0%; Pred. No. 3.3e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLLPKGVKHLKDFPILPGEI 25
|||
Db 484 RPLYSRRLLPKGVKHLKDFPILPGEI 508

RESULT 2
US-08-683-839B-3
; Sequence 3, Application US/08683839B
; Patent No. 5744326
; GENERAL INFORMATION:
; APPLICANT: Ill, Charles . R. et al.
; TITLE OF INVENTION: Use of Viral Cis-Acting Post-Transcriptional
; TITLE OF INVENTION: Regulatory Sequences To Increase Expression of
; TITLE OF INVENTION: Intronsless Genes Containing Near-Consensus Splice Sites
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1075

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/683,839B
FILING DATE: 11-MARCH-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Remillard, Jane E.
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: TTI-138
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1471 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-683-839B-3

Query Match 93.4%; Score 127; DB 1; Length 1471;
Best Local Similarity 96.0%; Pred. No. 3.4e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|| |||||
DB 503 RPLYSRRLPKGVKHLKDFPILPGEI 527

RESULT 3
US-08-882-083-2
Sequence 2, Application US/08882083
Patent No. 5869292
GENERAL INFORMATION:
APPLICANT: VOORBERG, Johannes J.
TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,083
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/558,107
FILING DATE: 13-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: ISACSON, John P.
REGISTRATION NUMBER: 33,715
REFERENCE/DOCKET NUMBER: 30472/212
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 1661 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-882-083-2
Query Match 93.4%; Score 127; DB 2; Length 1661;
Best Local Similarity 96.0%; Pred. No. 3.9e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|| |||||
DB 503 RPLYSRRLPKGVKHLKDFPILPGEI 527

RESULT 4
US-08-558-107-2
Sequence 2, Application US/08558107
Patent No. 5910481
GENERAL INFORMATION:
APPLICANT: VOORBERG, Johannes J.
TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/558,107
FILING DATE: 13-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ISACSON, John P.
REGISTRATION NUMBER: 33,715
REFERENCE/DOCKET NUMBER: 30472/212
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1661 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-558-107-2
Query Match 93.4%; Score 127; DB 2; Length 1661;
Best Local Similarity 96.0%; Pred. No. 3.9e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|| |||||
DB 503 RPLYSRRLPKGVKHLKDFPILPGEI 527

RESULT 5
US-09-243-539-2
Sequence 2, Application US/09243539
Patent No. 6130203
GENERAL INFORMATION:
APPLICANT: VOORBERG, Johannes J.
TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:

TISSUE TYPE: Liver cDNA sequence
US-08-251-937A-4

Query Match 93.4%; Score 127; DB 1; Length 2332;
Best Local Similarity 96.0%; Pred. No. 5.7e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 8

US-08-212-133A-2
Sequence 2, Application US/08212133A
Patent No. 5663060

GENERAL INFORMATION:
APPLICANT: Lollar, John S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 100 Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/212,133A
FILING DATE: March 11, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU/76677
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-572-6508
TELEFAX: 404-572-6555

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapien
TISSUE TYPE: Liver cDNA sequence
US-08-212-133A-2

Query Match 93.4%; Score 127; DB 1; Length 2332;
Best Local Similarity 96.0%; Pred. No. 5.7e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 9

US-08-276-594A-2

Sequence 2, Application US/08276594A
Patent No. 5693499
GENERAL INFORMATION:
APPLICANT: YONEMURA, Hiroshi
APPLICANT: TAJIMA, Yoshitaka
APPLICANT: SUGAWARA, Keishin
APPLICANT: MASUDA, Kenichi
TITLE OF INVENTION: PROCESS FOR PREPARING HUMAN COAGULATION
TITLE OF INVENTION: FACTOR VIII PROTEIN COMPLEX
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,594A
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/950,191
FILING DATE: 24-SEP-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 243262/1991
FILING DATE: 24-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: WEGNER, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/195/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-276-594A-2

Query Match 93.4%; Score 127; DB 1; Length 2332;
Best Local Similarity 96.0%; Pred. No. 5.7e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 10

US-08-474-503-2
Sequence 2, Application US/08474503
Patent No. 5744446

GENERAL INFORMATION:
APPLICANT: Emory University
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 1100 Peachtree Street, Suite 2800
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309
COMPUTER READABLE FORM:

COUNTRY: USA
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/037,601
FILING DATE: 26-JUN-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US94/13200
FILING DATE: 15-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,133
FILING DATE: 11-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 75-95F
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids

```
; EARLIER FILING DATE: 1994-11-15
; EARLIER APPLICATION NUMBER: U.S. 08/212,133
; EARLIER FILING DATE: 1994-03-11
; EARLIER APPLICATION NUMBER: U.S. 07/864,004
; EARLIER FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-179-2

Query Match          93.4%; Score 127; DB 4; Length 2332;
Best Local Similarity 96.0%; Pred. No. 5.7e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 15
US-09-523-656-2
; Sequence 2, Application US/09523656
; Patent No. 6458563
; GENERAL INFORMATION:
; APPLICANT: Lollar S., John
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-951
; CURRENT APPLICATION NUMBER: US/09/523,656
; CURRENT FILING DATE: 2000-03-10
; EARLIER APPLICATION NUMBER: 09/037,601
; EARLIER FILING DATE: 1998-03-10
; EARLIER APPLICATION NUMBER: 08/670,707
; EARLIER FILING DATE: 1996-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-523-656-2

Query Match          93.4%; Score 127; DB 4; Length 2332;
Best Local Similarity 96.0%; Pred. No. 5.7e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

Search completed: January 23, 2003, 09:04:50
Job time : 17 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 23, 2003, 09:02:53 ; Search time 11 Seconds
(without alignments)
45.860 Million cell updates/sec

Title: SEQ2-ASNAT3
Perfect score: 136
Sequence: 1 RPNYSRRLPKGVKHLKDFPILPGEI 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

20 42.5 31.2 2076 10 US-09-815-242-5815 Sequence 5815, Ap
21 42.5 31.2 2186 10 US-09-815-242-12913 Sequence 12913, A
22 42 30.9 91 10 US-09-864-761-40747 Sequence 40747, A
23 42 30.9 378 9 US-09-738-626-6570 Sequence 6570, Ap
24 42 30.9 748 10 US-09-949-192-43 Sequence 43, Appl
25 41.5 30.5 260 10 US-09-452-239-12 Sequence 12, Appl
26 41 30.1 11 9 US-09-984-245-259 Sequence 259, Appl
27 41 30.1 63 10 US-09-864-761-46749 Sequence 46749, A
28 41 30.1 119 9 US-09-738-626-6842 Sequence 6842, Ap
29 41 30.1 254 10 US-09-815-242-11506 Sequence 11506, A
30 41 30.1 275 10 US-09-220-091-1 Sequence 1, Appli
31 41 30.1 302 10 US-09-945-301-11 Sequence 11, Appl
32 41 30.1 343 10 US-09-802-853-4 Sequence 4, Appli
33 41 30.1 410 9 US-09-910-186A-32 Sequence 32, Appl
34 41 30.1 540 9 US-09-984-245-134 Sequence 134, App
35 41 30.1 848 9 US-10-011-588-45 Sequence 45, Appl
36 41 30.1 7968 9 US-10-077-130-5 Sequence 5, Appli
37 40.5 29.8 247 10 US-09-452-239-50 Sequence 50, Appl
38 40.5 29.8 258 10 US-09-452-239-4 Sequence 4, Appli
39 40.5 29.8 258 10 US-09-452-239-38 Sequence 38, Appl
40 40.5 29.8 258 10 US-09-452-239-49 Sequence 49, Appl
41 40.5 29.8 259 10 US-09-452-239-36 Sequence 36, Appl
42 40.5 29.8 259 10 US-09-452-239-42 Sequence 42, Appl
43 40.5 29.8 595 9 US-09-738-626-6447 Sequence 6447, Ap
44 40 29.4 71 10 US-09-764-860-335 Sequence 335, App
45 40 29.4 75 10 US-09-815-242-10459 Sequence 10459, A

ALIGNMENTS

RESULT 1
US-10-006-091-1
; Sequence 1, Application US/10006091
; Patent No. US20020102730A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255.1
; CURRENT APPLICATION NUMBER: US/10/006,091
; CURRENT FILING DATE: 2001-12-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-006-091-1

Query Match 93.4%; Score 127; DB 12; Length 1438;
Best Local Similarity 96.0%; Pred. No. 2.8e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|||
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 2
US-10-047-257-1
; Sequence 1, Application US/10047257
; Patent No. US20020115152A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	93.4	1438	12	US-10-006-091-1 Sequence 1, Appli
2	127	93.4	1438	12	US-10-047-257-1 Sequence 1, Appli
3	127	93.4	1471	12	US-10-095-718-2 Sequence 2, Appli
4	127	93.4	2332	9	US-09-957-641-2 Sequence 2, Appli
5	101	74.3	1431	12	US-10-095-718-4 Sequence 4, Appli
6	50	36.8	897	10	US-09-949-192-49 Sequence 49, Appli
7	48.5	35.7	405	10	US-09-863-475A-8 Sequence 8, Appli
8	48.5	35.7	432	10	US-09-733-524-7 Sequence 7, Appli
9	48.5	35.7	433	9	US-10-120-319-11 Sequence 11, Appli
10	47	34.6	345	9	US-09-738-626-6584 Sequence 5584, Ap
11	44.5	32.7	376	10	US-09-834-722-2 Sequence 2, Appli
12	44.5	32.7	441	9	US-09-738-626-4005 Sequence 4005, Ap
13	44	32.4	282	10	US-09-893-637-2 Sequence 2, Appli
14	44	32.4	282	10	US-09-970-711-21 Sequence 21, Appli
15	44	32.4	482	10	US-09-796-858-40 Sequence 40, Appli
16	44	32.4	491	10	US-09-997-664-3 Sequence 3, Appli
17	43	31.6	342	10	US-09-815-242-12104 Sequence 12104, A
18	42.5	31.2	1342	10	US-09-815-242-10438 Sequence 10438, A
19	42.5	31.2	1380	10	US-09-815-242-14043 Sequence 14043, A

```
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255.2
; CURRENT APPLICATION NUMBER: US/10/047,257
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-047-257-1

Query Match          93.4%; Score 127; DB 12; Length 1438;
Best Local Similarity 96.0%; Pred. No. 2.8e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 3
US-10-095-718-2
; Sequence 2, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; FILE REFERENCE: 35052/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1471
; TYPE: PRT
; ORGANISM: Homo sapiens B-domain deleted factor VIII
; FEATURE:
; OTHER INFORMATION: Homo sapiens BDO FVIII
US-10-095-718-2

Query Match          93.4%; Score 127; DB 12; Length 1471;
Best Local Similarity 96.0%; Pred. No. 2.8e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db 503 RPLYSRRLPKGVKHLKDFPILPGEI 527

RESULT 4
US-09-957-641-2
; Sequence 2, Application US/09957641
; Publication No. US20020182670A1
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-00
; CURRENT APPLICATION NUMBER: US/09/957,641
; CURRENT FILING DATE: 2001-09-16
```

```
; PRIOR APPLICATION NUMBER: US 60/234047
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/236460
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
US-09-957-641-2

Query Match          93.4%; Score 127; DB 9; Length 2332;
Best Local Similarity 96.0%; Pred. No. 4.8e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 5
US-10-095-718-4
; Sequence 4, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; FILE REFERENCE: 35052/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1431
; TYPE: PRT
; ORGANISM: canine B-domain deleted factor VIII
US-10-095-718-4

Query Match          74.3%; Score 101; DB 12; Length 1431;
Best Local Similarity 79.2%; Pred. No. 2.6e-07;
Matches 19; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PNYSSRRLPKGVKHLKDFPILPGEI 25
Db 498 PLHTGRLPKGVKHLKDMPIPLPGEI 521

RESULT 6
US-09-949-192-49
; Sequence 49, Application US/09949192
; Patent No. US20020142292A1
; GENERAL INFORMATION:
; APPLICANT: Parham, Christi L.
; APPLICANT: Gorman, Daniel L.
; APPLICANT: Kurata, Hirokazu
; APPLICANT: Arai, Naoko
; APPLICANT: Sana, Theodore R.
; APPLICANT: Mattson, Jeanine D.
; APPLICANT: Murphy, Erin E.
; APPLICANT: Savkoor, Chetan
; APPLICANT: Grein, Jeffery
; APPLICANT: Smith, Kathleen M.
```


APPLICANT: McElanahan, Terrill K.
TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS AND METHODS
FILE REFERENCE: DX01169K
CURRENT APPLICATION NUMBER: US/09/949,192
CURRENT FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: 60/231,267
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.1
SEQ ID NO 49
LENGTH: 897
TYPE: PRT
ORGANISM: Homo sapiens
US-09-949-192-49

Query Match 36.8%; Score 50; DB 10; Length 897;
Best Local Similarity 40.0%; Pred. No. 9.7;
Matches 12; Conservative 5; Mismatches 5; Indels 8; Gaps 2;

QY 1 RPNYSRRRLPKGVKHLKD-----FPILPGEI 25
|||: ||:| | ||: ||:
Db 379 RPYYSK---KGLEHKADLQHLFPVPPGHL 405

RESULT 7
US-09-863-475A-8
Sequence 8, Application US/09863475A
Patent No. US20020102688A1
GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS, GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
City: Arlington
State: Virginia
Country: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/863,475A
FILING DATE: 24-May-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/914,281
FILING DATE: 20-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 405 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-863-475A-8

Best Local Similarity 50.0%; Pred. No. 6.6;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 1 RPNYSRRRLPKGV-KHLKDFFP 19
| | | | : | | | : | | |
Db 319 RANYERFVPRGAFIHVDDEP 338

RESULT 8
US-09-733-524-7
Sequence 7, Application US/09733524
Patent No. US20020068347A1
GENERAL INFORMATION:
APPLICANT: The Governors of the University of Alberta, a Canada Corporation
APPLICANT: Taylor, Diane E.
APPLICANT: Ge, Zhongming
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING ALPHA-1,3
TITLE OF INVENTION: FUCOSYLTRANSFERASES AND EXPRESSION SYSTEMS FOR MAKING AND
TITLE OF INVENTION: EXPRESSING THEM
FILE REFERENCE: 07254/049002
CURRENT APPLICATION NUMBER: US/09/733,524
CURRENT FILING DATE: 2000-12-14
PRIOR APPLICATION NUMBER: 09/092,315
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/048,857
PRIOR FILING DATE: 1997-06-06
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 432
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: {0}...(0)
OTHER INFORMATION: Helicobacter pylori fucosyltransferase - Mfuct4
US-09-733-524-7

Query Match 35.7%; Score 48.5; DB 10; Length 432;
Best Local Similarity 50.0%; Pred. No. 7.1;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 1 RPNYSRRRLPKGV-KHLKDFFP 19
| | | | : | | | : | | |
Db 346 RANYERFVPRGAFIHVDDEP 365

RESULT 9
US-10-120-319-11
Sequence 11, Application US/10120319
Patent No. US20020164749A1
GENERAL INFORMATION:
APPLICANT: Taylor, Diane E.
APPLICANT: Ge, Zhongming
TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
FILE REFERENCE: 07254/049001
CURRENT APPLICATION NUMBER: US/10/120,319
CURRENT FILING DATE: 2002-04-09
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/092,315
PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/048,857
PRIOR FILING DATE: EARLIER FILING DATE: 1997-06-06
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 11
LENGTH: 433
TYPE: PRT
ORGANISM: Mus musculus
US-10-120-319-11

Query Match 35.7%; Score 48.5; DB 9; Length 433;
Best Local Similarity 50.0%; Pred. No. 7.1;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

Query Match 32.7%; Score 44.5; DB 10; Length 376;
Best Local Similarity 52.6%; Pred. NO. 25;
Matches 10; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

RESULT 13
US-09-893-637-2
; Sequence 2, Application US/09893637
; Patent No. US20020049313A1
; GENERAL INFORMATION:
; APPLICANT: COTTEN, MATTHEW
; BAKER, ADAM
; CHIOCCA, SUSANNA
; TITLE OF INVENTION: Method for Introducing Foreign Material into
; Higher Eukaryotic Cells
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., NW., SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/893,637
; FILING DATE: 29-Jun-2001

```

> CLASSIFICATION: <Unknown>
>
> PRIOR APPLICATION DATA:
>
> APPLICATION NUMBER: US 08/750,180
> FILING DATE: 14-FEB-1997
> APPLICATION NUMBER: PCT/EP95/01989
> FILING DATE: 26-MAY-1995
> APPLICATION NUMBER: DE P 44 18 825.0
> FILING DATE: 30-MAY-1994
> APPLICATION NUMBER: DE P 44 42 587.2
> FILING DATE: 30-NOV-1994
> ATTORNEY/AGENT INFORMATION:
>
> NAME: Peter A. Jackman
> REGISTRATION NUMBER: 45,986
> REFERENCE/DOCKET NUMBER: 0652.1580001
> TELECOMMUNICATION INFORMATION:
>
> TELEPHONE: 202-371-2600
> TELEFAX: 202-371-2540
>
> INFORMATION FOR SEQ ID NO: 2:
>
> SEQUENCE CHARACTERISTICS:
> LENGTH: 282 amino acids
> TYPE: amino acid
> TOPOLOGY: linear
>
> MOLECULE TYPE: Protein
>
> SEQUENCE DESCRIPTION: SEQ ID NO: 2:
>
> US-09-893-637-2

```

Query Match 32.4%; Score 44; DB 10; Length 282;
Best Local Similarity 52.6%; Pred. NO. 21;
Matches 10; Conservative 2; Mismatches 5; Indels 2; Gaps 1;

QY 4 YSRRLPKGVKHLKD--FPI 20
|| || | : : | |||
Db 124 YSDPLPHGITHVGD LGFPI 142

```

RESULT 14
US-09-970-711-21
; Sequence 21, Application US/09970711
; Patent NO. US20020081279A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Adam
; APPLICANT: Cotten, Matthew
; APPLICANT: Chiocca, Susanna
; APPLICANT: Kurzbaumer, Robert
; APPLICANT: Schaffner, Gotthold
; TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus
; FILE REFERENCE: 0652.1800001
; CURRENT APPLICATION NUMBER: US/09/970,711
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/171,461
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: PCT/EP97/01944
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 282
; TYPE: PRT
; ORGANISM: CELO Virus
; FEATURE:
; OTHER INFORMATION: Position: 37391..38239 /note=ORF8
US-09-970-711-21

```

Query Match	32.4%;	Score 44;	DB 10;	Length 282;
Best Local Similarity	52.6%;	Pred. NO. 21;		
Matches 10;	Conservative	2;	Mismatches	5; Indels 2; Gaps 1;

```

QY      4 YSRRLPKGVKHLKD--FPI 20
         || || || || || || || ||
Db     124 YSDPLPHGITHVGD LGFPI 142

```

RESULT 15

```

US-09-796-858-40
; Sequence 40, Application US/09796858
; Patent No. US20020055139A1
; GENERAL INFORMATION:
; APPLICANT: Holtzmann, Douglas
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING PROGNOSTIC, DIAGNOSTIC,
; TITLE OF INVENTION: PREVENTIVE, THERAPEUTIC, AND OTHER USES
; FILE REFERENCE: 7853-226-999
; CURRENT APPLICATION NUMBER: US/09/796,858
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 50
; SEQ ID NO 40
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-858-40

```

```

Query Match      32.4%; Score 44; DB 10; Length 482;
Best local Similarity 36.4%; Pred. NO. 39;
Matches 8; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 1 RPNYSRRRLPKGVKHLKDFPLP 22
    || : | : |||
db 450 RPSLQKHPSPTGGELHIAPFPLP 471

```

Search completed: January 23, 2003, 09:05:29
Job time : 12 secs

